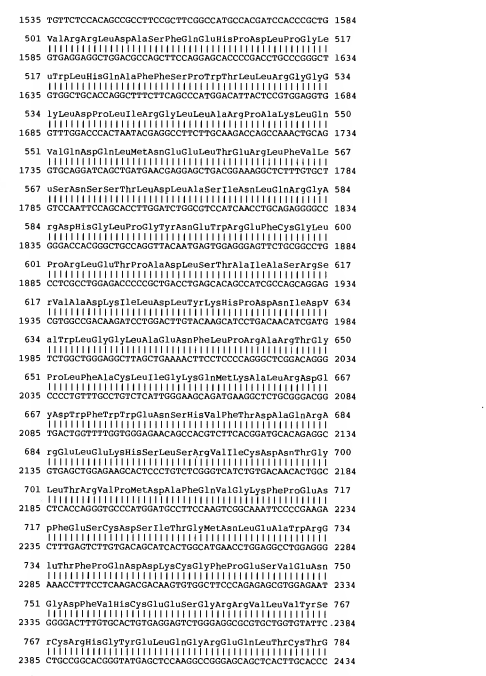
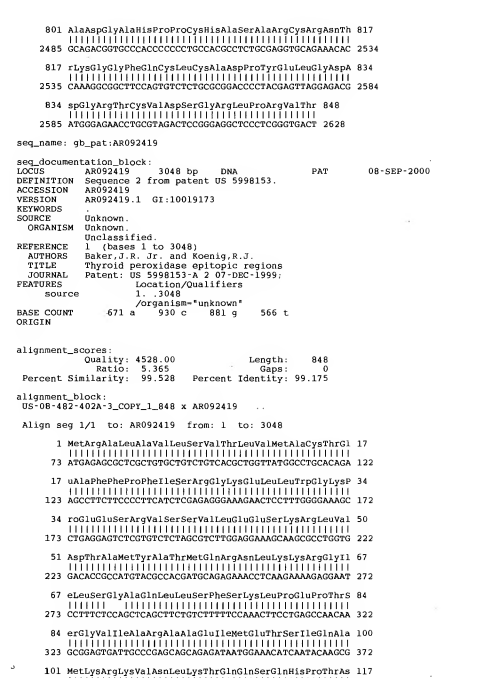
```
OM of: US-08-482-402A-3_COPY_1_848 to: GenEmbl:*   out_format : pfs

Date: Nov 26, 2001 11:24 AM

About: Results were produced by the GenCore software, version 4.5,
       Copyright (c) 1993-2000 Compugen Ltd.

Command line parameters:
 -MODEL=frame+_p2n.model -DEV=xlh
 -Q=/cgn2_1/USPTO_spool/US08482402/runat_26112001_091652_29053/app_query.fasta_1.931
 -DB=GenEmbl -QFMT=fastap -SUFFIX=rge -GAPOP=12.000 -GAPEXT=4.000
 -MINMATCH=0.100 -LOOPCL=0.000 -LOOPEXT=0.000 -QGAPOP=4.500
 -QGAPEXT=0.050 -XGAPOP=10.000 -XGAPEXT=0.500 -FGAPOP=6.000
 -FGAPEXT=7.000 -YGAPOP=10.000 -YGAPEXT=0.500 -DELOP=6.000
 -DELEXT=7.000 -START=1 -MATRIX=blosum62 -TRANS=human40.cdi
 -LIST=45 -DOCALIGN=200 -THR_SCORE=pct -THR_MAX=100 -THR_MIN=0
 -ALIGN=15 -MODE=LOCAL -OUTFMT=pfs -NORM=ext -HEAPSIZE=500
 -MINLEN=0 -MAXLEN=2000000000 -USER=US08482402_@CGN1_1_7160
 -NCPU=6 -ICPU=3 -LONGLOG -DEV_TIMEOUT=120 -WARN_TIMEOUT=30
 -NO_XLPXY -WAIT -THREADS=1

Search information block:
Query: US-08-482-402A-3_COPY_1_848
Query length: 848
Database: GenEmbl:*
Database sequences: 1472140
Database length: -341344837
Search time (sec): 1609.140000


score_list:
Sequence           Strd Orig    ZScore    EScore Len   ! Documentation  ..
gb_pr:HUMTPOC        + 4569.00 4424.87  3.4e-238  3060   ! M17755 Homo sapiens thyroid pe
gb_pat:AR092419      + 4528.00 4385.16  5.6e-236  3048   ! AR092419 Sequence 2 from pater
gb_pr:HUMTPOA        + 4528.00 4385.16  5.6e-236  3048   ! J02969 Human thyroid peroxidas
gb_pat:A92132        + 4525.00 4382.76  7.6e-236  2847   ! A92132 Sequence 1 from Patent
gb_pr:HSTPO          + 4525.00 4382.30  8.1e-236  3027   ! Y00406 Human mRNA for thyroper
gb_pat:E23825        + 4495.00 4354.51  2.8e-234  2546   ! E23825 Antigen for immunoassay
gb_pr:HSASTPO2       + 4215.50 4082.79  3.9e-219  2845   ! X17358 Human mRNA from alterna
gb_pr:HUMTPOB        + 4206.50 4073.99  1.2e-218  2877   ! J02970 Human thyroid peroxidas
gb_ro:RNTPO          + 3563.00 3449.43  7.4e-184  3237   ! X17396 Rat mRNA for thyroid pe
gb_ro:MMTPER         + 3520.00 3407.64  1.6e-181  3291   ! X60703 Mus musculus mRNA for t
gb_ro:RATTPOFR       + 3082.00 2984.38  5.9e-158  2777   ! M31655 Rat thyroid peroxidase
gb_in:AB022197       + 1927.50 1864.44  1.4e-95   3179   ! AB022197 Halocynthia roretzi m
gb_in:AB022196       + 1879.00 1818.07  5.4e-93   2915   ! AB022196 Ciona intestinalis mR
gb_ov:AF349034       + 1670.50 1613.94  1.3e-81   3846   ! AF349034 Danio rerio myeloperc
gb_ov:AF378824       + 1619.00 1566.34  5.7e-79   2814   ! AF378824 Danio rerio myeloid-s
gb_pat:A08802        + 1576.50 1526.77  9.1e-77   2261   ! A08802 H.sapiens gene for myel
gb_pr:HSMPOR         + 1576.50 1524.17  1.3e-76   3213   ! X04876 Human mRNA for myeloper
gb_pr:S56200         + 1575.50 1525.87  1.0e-76   2238   ! S56200 myeloperoxidase [human,
gb_pr:HUMMYP         + 1575.50 1524.85  1.2e-76   2568   ! J02694 Human myeloperoxidase m
gb_pat:E12629        + 1575.50 1523.19  1.4e-76   3215   ! E12629 cDNA encoding human nov
gb_pr:HUMMPO         + 1575.50 1523.19  1.4e-76   3215   ! M19507 Human myeloperoxidase m
gb_ro:MUSEPP         + 1564.50 1513.84  4.8e-76   2694   ! D78353 Mouse mRNA for eosinoph
gb_ro:MUSEPO         + 1559.50 1509.07  8.8e-76   2666   ! L77979 Mus musculus (clone pBM
gb_ro:MMMPO          + 1540.00 1489.97  1.0e-74   2740   ! X15313 Mouse MPO mRNA for myel
gb_in:AB028841       + 1532.00 1480.51  3.4e-74   3449   ! AB028841 Branchiostoma belcher
gb_pr:HSEP           + 1503.00 1454.62  9.6e-73   2558   ! X14346 Human mRNA for eosinoph
gb_in:AE003475       - 1438.50 1356.83  2.7e-67   300829  ! AE003475 Drosophila melanoga
gb_ov:XLU68724       + 1437.50 1391.27  3.2e-69   2512   ! U68724 Xenopus laevis polysoma
gb_htg:AC019531      - 1424.50 1353.04  4.3e-67   80254  ! AC019531 Drosophila melanogas
gb_pr:D86983         + 1421.00 1369.46  5.3e-68   5510   ! D86983 Human mRNA for KIAA023C
gb_pr:AF200348       + 1421.00 1367.86  6.5e-68   6847   ! AF200348 Homo sapiens melanoma
gb_pr:HSU39573       + 1409.50 1363.41  1.1e-67   2768   ! U39573 Human salivary peroxida
gb_om:BOVLPO         + 1404.00 1358.24  2.2e-67   2710   ! M58150 Bovine lactoperoxidase
gb_in:DMU11052       + 1391.50 1341.79  1.8e-66   4871   ! U11052 Drosophila melanogaster
gb_om:AF027970       + 1382.00 1336.73  3.5e-66   2780   ! AF027970 Ovis aries airway lac
gb_om:CDR131675      + 1379.50 1334.70  4.6e-66   2634   ! AJ131675 Camelus dromedarius m
gb_pr:AC009471       - 1137.50 1069.22  2.8e-51   172445  ! AC009471 Homo sapiens BAC cl
gb_pr:HUMTPOD        + 1019.00  992.14  5.5e-47   1046   ! M55702 Human thyroperoxidase m
gb_om:SSTPOR1        + 1007.00  979.86  2.7e-46   1142   ! X04645 Porcine mRNA for thyroi
gb_htg:AF186999      -  960.50  903.11  5.0e-42   82650  ! AF186999 Homo sapiens chromos
gb_htg:AC011128      +  945.50  886.74  4.1e-41   105849  ! AC011128 Homo sapiens chr
gb_ov:AF364820       +  939.50  912.68  1.5e-42   1448   ! AF364820 Xenopus laevis per
gb_in:CELZK994       +  930.50  882.72  6.8e-41   25547  ! AF022977 Caenorhabditis el
gb_htg:AC060811      +  926.50  865.24  6.4e-40   160674  ! AC060811 Homo sapiens chr
gb_pr:HUMTPO07       +  918.50  889.95  2.7e-41   1996   ! M25707 Human thyroid peroxi

seq_name: gb_pr:HUMTPOC

seq_documentation_block:
LOCUS       HUMTPOC      3060 bp    mRNA            PRI       26-APR-1999
DEFINITION  Homo sapiens thyroid peroxidase (TPO) mRNA, complete cds.
ACCESSION   M17755
VERSION     M17755.2  GI:4680720
KEYWORDS    .
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 732 to 1573)
  AUTHORS   Seto,P., Hirayu,H., Magnusson,R.P., Gestautas,J., Portmann,L.,
            DeGroot,L.J. and Rapoport,B.
  TITLE     Isolation of a complementary DNA clone for thyroid microsomal
            antigen. Homology with the gene for thyroid peroxidase
  JOURNAL   J. Clin. Invest. 80 (4), 1205-1208 (1987)
  MEDLINE   88008367
REFERENCE   2  (bases 1 to 3060)
  AUTHORS   Magnusson,R.P., Chazenbalk,G.D., Gestautas,J., Seto,P., Filetti,S.,
            DeGroot,L.J. and Rapoport,B.
  TITLE     Molecular cloning of the complementary deoxyribonucleic acid for
            human thyroid peroxidase
  JOURNAL   Mol. Endocrinol. 1 (11), 856-861 (1987)
  MEDLINE   91042572
REFERENCE   3  (bases 732 to 1573)
  AUTHORS   Rapoport,B.
  TITLE     Direct Submission
  JOURNAL   Submitted (03-AUG-1993) Autoimmune Disease Unit, Cedars-Sinai
            Research Institute, 8700 Beverly Blvd, Los Angeles, CA 90048, USA
REFERENCE   4  (bases 1 to 3060)
  AUTHORS   Rapoport,B.
  TITLE     Direct Submission
  JOURNAL   Submitted (26-APR-1999) Autoimmune Disease Unit, Cedars-Sinai
            Research Institute, 8700 Beverly Blvd, Los Angeles, CA 90048, USA
  REMARK    Sequence updated by submitter
COMMENT     On Apr 26, 1999 this sequence version replaced gi:339872.
FEATURES             Location/Qualifiers
     source          1..3060
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /chromosome="2"
                     /map="2p24"
                     /clone="PM-5"
                     /tissue_type="thyroid"
                     /note="isolated from patient with Graves' disease"
     gene            1..3060
                     /gene="TPO"
     CDS             85..2886
                     /gene="TPO"
                     /function="primary enzyme for thyroid hormone synthesis"
                     /note="similar to Homo sapiens thyroid peroxidases in
                     GenBank Accession Numbers J02969 and Y00406;
                     thyroperoxidase; membrane-associated cell surface protein;
                     primary autoantigen in human autoimmune thyroiditis
                     (Hashimoto's thyroiditis); thyroid microsomal antigen"
                     /codon_start=1
                     /product="thyroid peroxidase"
                     /protein_id="AAA61217.2"
                     /db_xref="GI:4680721"
                     /translation="MRALAVLSVTLVMACTEAFFPFISRGKELLWGKPEESRVSSVLE
                     ESKRLVDTAMYATMQRNLKKRGILSGAQLLSFSKLPEPTSGVIARAAEIMETSIQAMK
                     RKVNLKTQQSQHPTDALSEDLLSIIANMSGCLPYMLPPKCPNTCLANKYRPITGACNN
                     RDHPRWGASNTALARWLPPVYEDGFSQPRGWNPGFLYNGFPLPPVREVTRHVIQVSNE
                     VVTDDDRYSDLLMAWGQYIDHDIAFTPQSTSKAAFGGGSDCQMTCENQNPCFPIQLPE
                     EARPAAGTACLPFYRSSAACGTGDQGALFGNLSTANPRQQMNGLTSFLDASTVYGSSP
```

```
                    ALERQLRNWTSAEGLLRVHGRLRDSGRAYLPFVPPRAPAACAPEPGNPGETRGPCFLA
                    GDGRASEVPSLTALHTLWLREHNRLAAALKALNAHWSADAVYQEARKVVGALHQIITL
                    RDYIPRILGPEAFQQYVGPYEGYDSTANPTVSNVFSTAAFRFGHATIHPLVRRLDASF
                    QEHPDLPGLWLHQAFFSPWTLLRGGGLDPLIRGLLARPAKLQVQDQLMNEELTERLFV
                    LSNSSTLDLASINLQRGRDHGLPGYNEWREFCGLPRLETPADLSTAIASRSVADKILD
                    LYKHPDNIDVWLGGLAENFLPRARTGPLFACLIGKQMKALRDGDWFWWENSHVFTDAQ
                    RRELEKHSLSRVICDNTGLTRVPMDAFQVGKFPEDFESCDSITGMNLEAWRETFPQDD
                    KCGFPESVENGDFVHCEESGRRVLVYSCRHGYELQGREQLTCTQEGWDFQPPLCKDVN
                    ECADGAHPPCHASARCRNTKGGFQCLCADPYELGDDGRTCVDSGRLPRVTWISMSLAA
                    LLIGGFAGLTSTVICRWTRTGTKSTLPISETGGGTPELRCGKHQAVGTSPQRAAAQDS
                    EQESAGMEGRDTHRLPRAL"
     misc_feature   2626..2697
                    /gene="TPO"
                    /note="putative; transmembrane-region site"
BASE COUNT      677 a    930 c    887 g    566 t
ORIGIN


alignment_scores:
           Quality: 4569.00              Length:     848
             Ratio:   5.388                Gaps:       0
 Percent Similarity: 100.000   Percent Identity: 100.000

alignment_block:
 US-08-482-402A-3_COPY_1_848 x HUMTPOC   ..

 Align seg 1/1  to: HUMTPOC  from: 1  to: 3060

       1 MetArgAlaLeuAlaValLeuSerValThrLeuValMetAlaCysThrGl 17
         ||||||||||||||||||||||||||||||||||||||||||||||||||
      85 ATGAGAGCGCTGGCTGTGCTGTCTGTCACGCTGGTTATGGCCTGCACAGA 134

      17 uAlaPhePheProPheIleSerArgGlyLysGluLeuLeuTrpGlyLysP 34
         ||||||||||||||||||||||||||||||||||||||||||||||||||
     135 AGCCTTCTTCCCCTTCATCTCGAGAGGGAAAGAACTCCTTTGGGGAAAGC 184

      34 roGluGluSerArgValSerSerValLeuGluGluSerLysArgLeuVal 50
         ||||||||||||||||||||||||||||||||||||||||||||||||||
     185 CTGAGGAGTCTCGTGTCTCTAGCGTCTTGGAGGAAAGCAAGCGCCTGGTG 234

      51 AspThrAlaMetTyrAlaThrMetGlnArgAsnLeuLysLysArgGlyIl 67
         ||||||||||||||||||||||||||||||||||||||||||||||||||
     235 GACACCGCCATGTACGCCACGATGCAGAGAAACCTCAAGAAAAGAGGAAT 284

      67 eLeuSerGlyAlaGlnLeuLeuSerPheSerLysLeuProGluProThrS 84
         ||||||||||||||||||||||||||||||||||||||||||||||||||
     285 CCTTTCTGGAGCTCAGCTTCTGTCTTTTTCCAAACTTCCTGAGCCAACAA 334

      84 erGlyValIleAlaArgAlaAlaGluIleMetGluThrSerIleGlnAla 100
         ||||||||||||||||||||||||||||||||||||||||||||||||||
     335 GCGGAGTGATTGCCCGAGCAGCAGAGATAATGGAAACATCAATACAAGCG 384

     101 MetLysArgLysValAsnLeuLysThrGlnGlnSerGlnHisProThrAs 117
         ||||||||||||||||||||||||||||||||||||||||||||||||||
     385 ATGAAAAGAAAAGTCAACCTGAAAACTCAACAATCACAGCATCCAACGGA 434

     117 pAlaLeuSerGluAspLeuLeuSerIleIleAlaAsnMetSerGlyCysL 134
         ||||||||||||||||||||||||||||||||||||||||||||||||||
     435 TGCTTTATCAGAAGATCTGCTGAGCATCATTGCAAACATGTCTGGATGTC 484

     134 euProTyrMetLeuProProLysCysProAsnThrCysLeuAlaAsnLys 150
         ||||||||||||||||||||||||||||||||||||||||||||||||||
     485 TCCCTTACATGCTGCCCCCAAAATGCCCAAACACTTGCCTGGCGAACAAA 534

     151 TyrArgProIleThrGlyAlaCysAsnAsnArgAspHisProArgTrpGl 167
         ||||||||||||||||||||||||||||||||||||||||||||||||||
     535 TACAGGCCCATCACAGGAGCTTGCAACAACAGAGACCACCCCAGATGGGG 584

     167 yAlaSerAsnThrAlaLeuAlaArgTrpLeuProProValTyrGluAspG 184
         ||||||||||||||||||||||||||||||||||||||||||||||||||
     585 CGCCTCCAACACGGCCCTGGCACGATGGCTCCCTCCAGTCTATGAGGACG 634

     184 lyPheSerGlnProArgGlyTrpAsnProGlyPheLeuTyrAsnGlyPhe 200
         ||||||||||||||||||||||||||||||||||||||||||||||||||
     635 GCTTCAGTCAGCCCCGAGGCTGGAACCCCGGCTTCTTGTACAACGGGTTC 684

     201 ProLeuProProValArgGluValThrArgHisValIleGlnValSerAs 217
         ||||||||||||||||||||||||||||||||||||||||||||||||||
     685 CCACTGCCCCCGGTCCGGGAGGTGACAAGACATGTCATTCAAGTTTCAAA 734

     217 nGluValValThrAspAspAspArgTyrSerAspLeuLeuMetAlaTrpG 234
         ||||||||||||||||||||||||||||||||||||||||||||||||||
     735 TGAGGTTGTCACAGATGATGACCGCTATTCTGACCTCCTGATGGCATGGG 784

     234 lyGlnTyrIleAspHisAspIleAlaPheThrProGlnSerThrSerLys 250
         ||||||||||||||||||||||||||||||||||||||||||||||||||
     785 GACAATACATCGACCACGACATCGCGTTCACACCACAGAGCACCAGCAAA 834

     251 AlaAlaPheGlyGlyGlySerAspCysGlnMetThrCysGluAsnGlnAs 267
         ||||||||||||||||||||||||||||||||||||||||||||||||||
     835 GCTGCCTTCGGGGGAGGGTCTGACTGCCAGATGACTTGTGAGAACCAAAA 884

     267 nProCysPheProIleGlnLeuProGluGluAlaArgProAlaAlaGlyT 284
         ||||||||||||||||||||||||||||||||||||||||||||||||||
     885 CCCATGTTTTCCCATACAACTCCCGGAGGAGGCCCGGCCGGCCGCGGGCA 934

     284 hrAlaCysLeuProPheTyrArgSerSerAlaAlaCysGlyThrGlyAsp 300
         ||||||||||||||||||||||||||||||||||||||||||||||||||
     935 CCGCCTGTCTGCCCTTCTACCGCTCTTCGGCCGCCTGCGGCACCGGGGAC 984

     301 GlnGlyAlaLeuPheGlyAsnLeuSerThrAlaAsnProArgGlnGlnMe 317
         ||||||||||||||||||||||||||||||||||||||||||||||||||
     985 CAAGGCGCGCTCTTTGGGAACCTGTCCACGGCCAACCCGAGGCAGCAGAT 1034

     317 tAsnGlyLeuThrSerPheLeuAspAlaSerThrValTyrGlySerSerP 334
         ||||||||||||||||||||||||||||||||||||||||||||||||||
    1035 GAACGGGTTGACCTCGTTCCTGGACGCGTCCACCGTGTATGGCAGCTCCC 1084

     334 roAlaLeuGluArgGlnLeuArgAsnTrpThrSerAlaGluGlyLeuLeu 350
         ||||||||||||||||||||||||||||||||||||||||||||||||||
    1085 CGGCCCTAGAGAGGCAGCTGCGGAACTGGACCAGTGCCGAAGGGCTGCTC 1134

     351 ArgValHisGlyArgLeuArgAspSerGlyArgAlaTyrLeuProPheVa 367
         ||||||||||||||||||||||||||||||||||||||||||||||||||
    1135 CGCGTCCACGGCCGCCTCCGGGACTCCGGCCGCGCCTACCTGCCCTTCGT 1184

     367 lProProArgAlaProAlaAlaCysAlaProGluProGlyAsnProGlyG 384
         ||||||||||||||||||||||||||||||||||||||||||||||||||
    1185 GCCGCCACGCGCGCCTGCGGCCTGTGCGCCCGAGCCCGGCAACCCCGGAG 1234

     384 luThrArgGlyProCysPheLeuAlaGlyAspGlyArgAlaSerGluVal 400
         ||||||||||||||||||||||||||||||||||||||||||||||||||
    1235 AGACCCGCGGGCCCTGCTTCCTGGCCGGAGACGGCCGCGCCAGCGAGGTC 1284

     401 ProSerLeuThrAlaLeuHisThrLeuTrpLeuArgGluHisAsnArgLe 417
         ||||||||||||||||||||||||||||||||||||||||||||||||||
    1285 CCCTCCCTGACGGCACTGCACACGCTGTGGCTGCGCGAGCACAACCGCCT 1334

     417 uAlaAlaAlaLeuLysAlaLeuAsnAlaHisTrpSerAlaAspAlaValT 434
         ||||||||||||||||||||||||||||||||||||||||||||||||||
    1335 GGCCGCGGCGCTCAAGGCCCTCAATGCGCACTGGAGCGCGGACGCCGTGT 1384

     434 yrGlnGluAlaArgLysValValGlyAlaLeuHisGlnIleIleThrLeu 450
         ||||||||||||||||||||||||||||||||||||||||||||||||||
    1385 ACCAGGAGGCGCGCAAGGTCGTGGGCGCTCTGCACCAGATCATCACCCTG 1434

     451 ArgAspTyrIleProArgIleLeuGlyProGluAlaPheGlnGlnTyrVa 467
         ||||||||||||||||||||||||||||||||||||||||||||||||||
    1435 AGGGATTACATCCCCAGGATCCTGGGACCCGAGGCCTTCCAGCAGTACGT 1484

     467 lGlyProTyrGluGlyTyrAspSerThrAlaAsnProThrValSerAsnV 484
         ||||||||||||||||||||||||||||||||||||||||||||||||||
    1485 GGGTCCCTATGAAGGCTATGACTCCACCGCCAACCCCACTGTGTCCAACG 1534

     484 alPheSerThrAlaAlaPheArgPheGlyHisAlaThrIleHisProLeu 500
         ||||||||||||||||||||||||||||||||||||||||||||||||||
```


1535 TGTTCTCCACAGCCGCCTTCCGCTTCGGCCATGCCACGATCCACCCGCTG 1584
501 ValArgArgLeuAspAlaSerPheGlnGluHisProAspLeuProGlyLe 517
1585 GTGAGGAGGCTGGACGCCAGCTTCCAGGAGCACCCCGACCTGCCCGGGCT 1634
517 uTrpLeuHisGlnAlaPheSerProTrpThrLeuLeuArgGlyGlyG 534
1635 GTGGCTGCACCAGGCTTCTTCAGCCCATGGACATTACTCCGTGGAGGTG 1684
534 lyLeuAspProLeuIleArgGlyLeuLeuAlaArgProAlaLysLeuGln 550
1685 GTTTGGACCCACTAATACGAGGCCTTCTTGCAAGACCAGCCAAACTGCAG 1734
551 ValGlnAspGlnLeuMetAsnGluGluLeuThrGluArgLeuPheValLe 567
1735 GTGCAGGATCAGCTGATGAACGAGGAGCTGACGGAAAGGCTCTTTGTGCT 1784
567 uSerAsnSerSerThrLeuAspLeuAlaSerIleAsnLeuGlnArgGlyA 584
1785 GTCCAATTCCAGCACCTTGGATCTGGCTCCATCAACCTGCAGAGGGGCC 1834
584 rgAspHisGlyLeuProGlyTyrAsnGluTrpArgGluPheCysGlyLeu 600
1835 GGGACCACGGCTGCCAGGTTACAATGAGTGGAGGAGTTCTGCGGCCTG 1884
601 ProArgLeuGluThrProAlaAspLeuSerThrAlaIleAlaSerArgSe 617
1885 CCTCGCCTGGAGACCCCCGCTGACCTGAGCACAGCCATCGCCAGCAGGAG 1934
617 rValAlaAspLysIleLeuAspLeuTyrLysHisProAspAsnIleAspV 634
1935 CGTGGCGCAGAAGATCCTGGACTTGTACAAGCATCCTGACAACATCGATG 1984
634 alTrpLeuGlyGlyLeuAlaGluAsnPheLeuProArgAlaArgThrGly 650
1985 TCTGGCTGGGAGGCTTAGCTGAAAACTTCCTCCCCAGGGCTCGGACAGGG 2034
651 ProLeuPheAlaCysLeuIleGlyLysGlnMetLysAlaLeuArgAspGl 667
2035 CCCCTGTTTGCCTGTCTCATTGGGAAGCAGATGAAGGCTCTGCGGACGG 2084
667 yAspTrpPheTrpTrpGluAsnSerHisValPheThrAspAlaGlnArgA 684
2085 TGACTGGTTTTGGTGGGAGAACAGCCACGTCTTCACGGATGCACAGAGGC 2134
684 rgGluLeuGluLysHisSerLeuSerArgValIleCysAspAsnThrGly 700
2135 GTGAGCTGGAGAAGCACTCCCTGTCTCGGGTCATCTGTGACAACACTGGC 2184
701 LeuThrArgValProMetAspAlaPheGlnValGlyLysPheProGluAs 717
2185 CTCACCAGGGTGCCCATGGATGCCTTCCAAGTCGGCAAATTCCCCGAAGA 2234
717 pPheGluSerCysAspSerIleThrGlyMetAsnLeuGluAlaTrpArgG 734
2235 CTTTGAGTCTTGTGACAGCATCACTGGCATGAACCTGGAGGCCTGGAGGG 2284
734 luThrPheProGlnAspAspLysCysGlyPheProGluSerValGluAsn 750
2285 AAACCTTTCCTCAAGACGACAAGTGTGGCTTCCCAGAGAGCGTGGAGAAT 2334
751 GlyAspPheValHisCysGluGluSerGlyArgArgValLeuValTyrSe 767
2335 GGGGACTTTGTGCACTGTGAGGAGTCTGGGAGGCGCGTGCTGGTGTATTC 2384
767 rCysArgHisGlyTyrGluLeuGlnGlyArgGluGlnLeuThrCysThrG 784
2385 CTGCCGCACGGGTATGAGCTCCAAGCCGGGAGCAGCTCACTGCACCC 2434
784 lnGluGlyTrpAspPheGlnProProLeuCysLysAspValAsnGluCys 800
2435 AGGAAGGATGGGATTTCCAGCCTCCCCTCTGCAAAGATGTGAACGAGTGT 2484


801 AlaAspGlyAlaHisProProCysHisAlaSerAlaArgCysArgAsnTh 817
2485 GCAGACGGTGCCCACCCCCCCTGCCACGCCTCTGCGAGGTGCAGAAACAC 2534
817 rLysGlyGlyPheGlnCysLeuCysAlaAspProTyrGluLeuGlyAspA 834
2535 CAAAGGCGGCTTCCAGTGTCTCTGCGCGGACCCCTACGAGTTAGGAGACG 2584
834 spGlyArgThrCysValAspSerGlyArgLeuProArgValThr 848
2585 ATGGGAGAACCTGCGTAGACTCCGGGAGGCTCCCTCGGGTGACT 2628

seq_name: gb_pat:AR092419

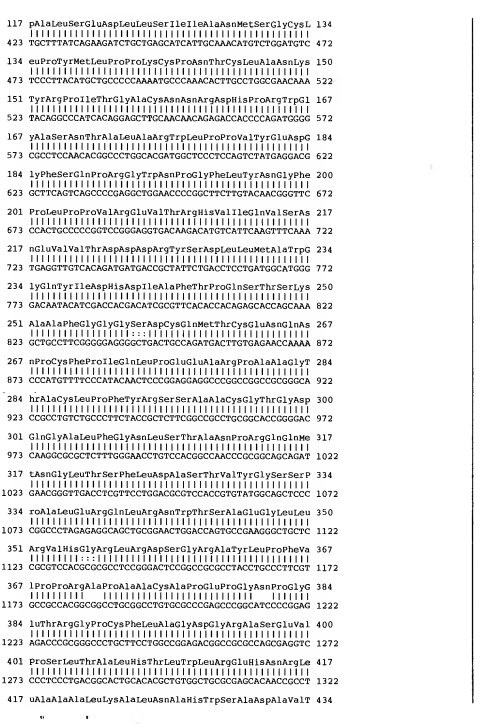
seq_documentation_block:
LOCUS AR092419 3048 bp DNA PAT 08-SEP-2000
DEFINITION Sequence 2 from patent US 5998153.
ACCESSION AR092419
VERSION AR092419.1 GI:10019173
KEYWORDS .
SOURCE Unknown.
ORGANISM Unknown.
Unclassified.
REFERENCE 1 (bases 1 to 3048)
AUTHORS Baker,J.R. Jr. and Koenig,R.J.
TITLE Thyroid peroxidase epitopic regions
JOURNAL Patent: US 5998153-A 2 07-DEC-1999;
FEATURES Location/Qualifiers
source 1..3048
/organism="unknown"
BASE COUNT 671 a 930 c 881 g 566 t
ORIGIN

alignment_scores:
Quality: 4528.00 Length: 848
Ratio: 5.365 Gaps: 0
Percent Similarity: 99.528 Percent Identity: 99.175

alignment_block:
US-08-482-402A-3_COPY_1_848 x AR092419 ..

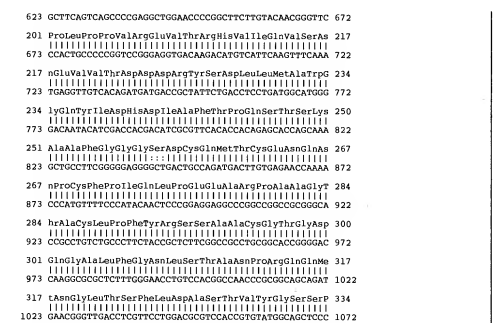
Align seg 1/1 to: AR092419 from: 1 to: 3048

1 MetArgAlaLeuAlaValLeuSerValThrLeuValMetAlaCysThrGl 17
73 ATGAGAGCGCTCGCTGTGCTGTCTGTCACGCTGGTTATGCCCTGCACAGA 122
17 uAlaPhePheProPheIleSerArgGlyLysGluLeuLeuTrpGlyLysP 34
123 AGCCTTTCCCCCTTCATCTCCGAGAGGGAAAGAACTCCTTTGGGGAAAGC 172
34 roGluGluSerArgValSerSerValLeuGluGluSerLysArgLeuVal 50
173 CTGAGGAGTCTCGTGTCTCTAGCGTCTTGGAGGAAAGCAAGCGCCTGGTG 222
51 AspThrAlaMetTyrAlaThrMetGlnArgAsnLeuLysLysArgGlyI1 67
223 GACACCGCCATGTACGCCACGATGCAGAGAAACCTCAAGAAAAGAGGAAT 272
67 eLeuSerGlyAlaGlnLeuLeuSerPheSerLysLeuProGluProThrS 84
273 CCTTTCTCCAGCTCAGCTTCTGTCTTTTCCAAACTTCCTGAGCCAACAA 322
84 erGlyValIleAlaArgAlaAlaGluIleMetGluThrSerIleGlnAla 100
323 GCGGAGTGATTGCCCGAGCAGCAGAGATAATGGAAACATCAATACAAGCG 372
101 MetLysArgLysValAsnLeuLysThrGlnGlnSerGlnHisProThrAs 117
373 ATGAAAAGAAAAGTCAACCTGAAAACTCAACAATCACAGCATCCAACGGA 422

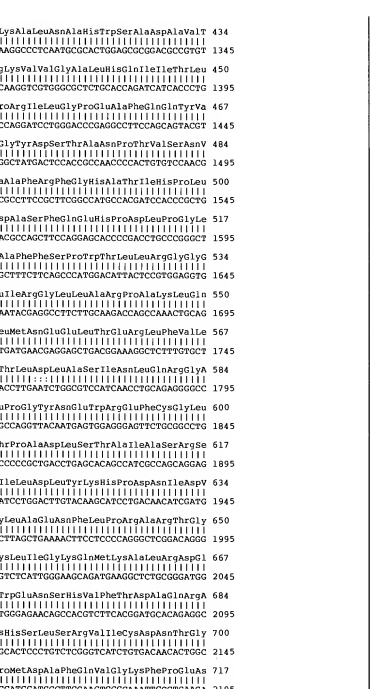

117 pAlaLeuSerGluAspLeuLeuSerIleIleAlaAsnMetSerGlyCysL 134
423 TGCTTTATCAGAAGATCTGCTGAGCATCATTGCAAACATGTCTGGATGTC 472
134 euProYrMetLeuProProLysCysProAsnThrCysLeuAlaAsnLys 150
473 TCCCTTACATGCTGCCCCCAAAATGCCCAAACACTTGCCTGCGAACAAA 522
151 TyrArgProIleThrGlyAlaCysAsnAsnArgAspHisProArgTrpGl 167
523 TACAGGCCATCACAGGAGCTTGCAACAACAGACGACCACCCAGATGGGG 572
167 yAlaSerAsnThrAlaLeuAlaArgTrpLeuProProValTyrGluAspG 184
573 CGCCTCCAACACGGCCTGGCACGATGGCTCCCTCCAGTCTATGAGGACG 622
184 lyPheSerGlnProArgGlyTrpAsnProGlyPheLeuTyrAsnGlyPhe 200
623 GCTTCAGTCAGCCCGAGGCTGGAACCCCGGCTTCTTGTACAACGGGTTC 672
201 ProLeuProProValArgGluValThrArgHisValIleGlnValSerAs 217
673 CCACTGCCCCCGGTCCGGGAGGTGACAAGACATGTCATTCAAGTTTCAAA 722
217 nGluValValThrAspAspAspArgTyrSerAspLeuMetAlaTrpG 234
723 TGAGGTTGTCACAGATGACGACGCTATTCTACTCGATGACATGGG 772
234 lyGlnTyrIleAspHisAspIleAlaPheThrProGlnSerThrSerLys 250
773 GACAATACATCGACCACGACATCGCGTTCACACCACAGAGCACCAGCAAA 822
251 AlaAlaPheGlyGlyGlySerAspCysGlnMetThrCysGluAsnGlnAs 267
823 GCTGCCTTCGGGGGAGGGGCTGACTGCCAGATGACTTGTGAGAACCAAAA 872
267 nProCysPheProIleGlnLeuProGluGluAlaArgProAlaAlaglyT 284
873 CCCATGTTTTCCCATACAACTCCCGGAGGAGGCCCGGCCGGCCGGGGCA 922
284 hrAlaCysLeuProPheTyrArgSerSerAlaAlaCysGlyThrGlyAsp 300
923 CCGCCTGTCTGCCCTTCTACCGCTCTTCGGCCGCCTGCGGCACCGGGGAC 972
301 GlnGlyAlaLeuPheGlyAsnLeuSerThrAlaAsnProArgGlnGlnMe 317
973 CAAGGCGCGCTCTTTGGGAACCTGTCCACGGCCAACCCGCGGCAGCAGAT 1022
317 tAsnGlyLeuThrSerPheLeuAspAlaSerThrValTyrGlySerSerP 334
1023 GAACGGTTGACCTCGTTCCTGGAGCGCGTCCACCGTGTATGGCAGCTCCC 1072
334 roAlaLeuGluArgGlnLeuArgAsnTrpThrSerAlaGluGlyLeuLeu 350
1073 CGGCCCTAGAGAGGCAGCTGCGGAACTGGACCAGTGCCGAAGGGCTGCTC 1122
351 ArgValHisGlyArgLeuArgAspSerGlyArgAlaTyrLeuProPheVa 367
1123 CGCGTCCACGCCGCCCTCCGGGACTCCGGCCGCGCCTACCTGCCCTTCGT 1172
367 lProProArgAlaProAlaAlaCysAlaProGluProGlyAsnProGlyG 384
1173 GCCGCCACGGCGGCCTGCGGCCTGTCGCCCGAGCCCGGCATCCCCGGAG 1222
384 lvThrArgGlyProCysPheLeuAlaGlyAspGlyArgAlaSerGluVal 400
1223 AGACCCGCGGGCCCTGCTTCCTGGCCGGAGACGGCCGCGCCAGCGAGGTC 1272
401 ProSerLeuThrAlaLeuHisThrLeuTrpLeuArgGluHisAsnArgLe 417
1273 CCCTCCCTGACGGCACTGCACACGCTGTGGCTGCGCGAGCACAACCGCCT 1322
417 uAlaAlaAlaLeuLysAlaLeuAsnAlaHisTrpSerAlaAspAlaValT 434
1323 GGCCGCGGCGCTCAAGGCCCTCAATGCGCACTGGAGCGCGGACGCCGTGT 1372
434 yrGlnGluAlaArgLysValValGlyAlaLeuHisGlnIleIleThrLeu 450
1373 ACCAGGAGGCGCCGAAGGTCGTGGCGCCTCTGCACCAGATCATCACCCTG 1422
451 ArgAspTyrIleProArgIleLeuGlyProGluAlaPheGlnGlnTyrVa 467
1423 AGGGATTACATCCCGAGGATCCTGGGACCCGAGGCCTTCCAGCAGTACGT 1472
467 lGlyProTyrGluGlyTyrAspSerThrAlaAsnProThrValSerAsnV 484
1473 GGGTCCCTATGAAGGCTATGACTCCACCGCCAACCCCACTGTGTCCAACG 1522
484 alPheSerThrAlaAlaPheArgPheGlyHisAlaThrIleHisProLeu 500
1523 TGTTCCTCACAGCCGCCTTCCGCTTCGGCTCGGCCATGCCACGATCCACCCGCTG 1572
501 ValArgArgLeuAspAlaSerPheGlnGluHisProAspLeuProGlyLe 517
1573 GTGAGGAGGCTGGACGCCAGCTTCCAGGAGCACCCCGACCTGCCCGGGCT 1622
517 uTrpLeuHisGlnAlaPhePheSerProTrpThrPheLeuArgGlyGlyG 534
1623 GTGGCTGCACCAGGCTTTCTTCAGCCCATGGACATTACTCCGTGGAGGTG 1672
534 lyLeuAspProLeuIleArgGlyLeuLeuAlaArgProAlaLysLeuGln 550
1673 GTTTGGACCCACTAATACGAGGCCTTCTCGCAAGACCAGCCAAACTGCAG 1722
551 ValGlnAspGlnLeuMetAsnGluGluLeuThrGluArgLeuPheValLe 567
1723 GTGCAGGATCAGCTGATGAACGAGGAGCTGACGGAAAGGCTCTTTGTGCT 1772
567 uSerAsnSerSerThrLeuAspLeuAlaSerIleAsnLeuGlnArgGlyA 584
1773 GTCCAATTCCAGCACCTTGGATCTGGCGTCCATCAACCTGCAGAGGGGCC 1822
584 rgAspHisGlyLeuProGlyTyrAsnGluTrpArgGluPheCysGlyLeu 600
1823 GGGACCACGGGCTGCCAGGTTACAATGATGGAGGAGGTTCTGCGGCCTG 1872
601 ProArgLeuGluThrProAlaAspLeuSerThrAlaIleAlaSerArgSe 617
1873 CCTCGCCTGGAGACCCCCGCTGACCTGAGCACAGCCATCGCAGCAGGAG 1922
617 rValAlaAspLysIleLeuAspLeuTyrLysHisProAspAsnIleAspV 634
1923 CGTGGCCGACAAGATCCTGGACTTGTACAAGCATCCTGACAACATCGATG 1972
634 alTrpLeuGlyGlyLeuAlaGluAsnPheLeuProArgAlaArgThrGly 650
1973 TCTGGCTGGGAGCTTAGCTGAAAACTTCCTCCCCAGGGCTCGGACAGGG 2022
651 ProLeuPheAlaCysLeuIleGlyLysGlnMetLysAlaLeuArgAspGl 667
2023 CCCCTGTTTGCCTGTCTCATTGGGAAGCAGATGAAGGCTCTGCGGGATGG 2072
667 yAspTrpPheTrpTrpGluAsnSerHisValPheThrAspAlaGlnArgA 684
2073 TGACTGGTTTTGGTGGGAGAACAGCCACGTCTTCACGGATGCACAGAGGC 2122
684 rgGluLeuGluLysHisSerLeuSerArgValIleCysAspAsnThrGly 700
2123 GTGAGCTGGAGAAGCACTCCCTGTCTCGGGTCATCTGTGACAACACTGGC 2172
701 LeuThrArgValProMetAspAlaPheGlnValGlyLysPheProGluAs 717
2173 CTCACCAGGGTGCCCATGGATGCCTTCCAAGTCGGCAAATTCCCTGAAGA 2222
717 pPheGluSerCysAspSerIleThrGlyMetAsnLeuGluAlaTrpArgG 734

```
  2223 CTTTGAGTCTTGTGACAGCATCCCTGGCATGAACCTGGAGGCCTGGAGGG 2272

   734 luThrPheProGlnAspAspLysCysGlyPheProGluSerValGluAsn 750
       ||||||||||||||||||||||||||||||||||||||||||||||||||
  2273 AAACCTTTCCTCAAGACGACAAGTGTGGCTTCCCAGAGAGCGTGGAGAAT 2322

   751 GlyAspPheValHisCysGluGluSerGlyArgArgValLeuValTyrSe 767
       ||||||||||||||||||||||||||||||||||||||||||||||||||
  2323 GGGGCATTTGTGCACTGTGAGGAGTCTGGGAGGCGCGTGCTGGTGTATTC 2372

   767 rCysArgHisGlyTyrGluLeuGlnGlyArgGluGlnLeuThrCysThrG 784
       ||||||||||||||||||||||||||||||||||||||||||||||||||
  2373 CTGCCGGCACGGGTATGAGCTCCAAGGCCGGGAGCAGCTCACTTGCACCC 2422

   784 lnGluGlyTrpAspPheGlnProProLeuCysLysAspValAsnGluCys 800
       ||||||||||||||||||||||||||||||||||||||||||||||||||
  2423 AGGAAGGATGGGATTTCCAGCCTCCCCTCTGCAAAGATGTGAACGAGTGT 2472

   801 AlaAspGlyAlaHisProProCysHisAlaSerAlaArgCysArgAsnTh 817
       ||||||||||||||||||||||||||||||||||||||||||||||||||
  2473 GCAGACGGTGCCCACCCCCCCTGCCACGCCTCTGCGAGGTGCAGAAACAC 2522

   817 rLysGlyGlyPheGlnCysLeuCysAlaAspProTyrGluLeuGlyAspA 834
       ||||||||||||||||||||||||||||||||||||||||||||||||||
  2523 CAAAGGCGGCTTCCAGTGTCTCTGCGCGGACCCCTACGAGTTAGGAGACG 2572

   834 spGlyArgThrCysValAspSerGlyArgLeuProArgValThr 848
       ||||||||||||||||||||||||||||||||||||||||||||
  2573 ATGGGAGAACCTGCGTAGACTCCGGGAGGCTCCCTCGGGCGACT 2616

seq_name: gb_pr:HUMTPOA

seq_documentation_block:
LOCUS       HUMTPOA      3048 bp    mRNA            PRI      14-JAN-1995
DEFINITION  Human thyroid peroxidase mRNA, clone phTPO-2.8.
ACCESSION   J02969
VERSION     J02969.1  GI:339866
KEYWORDS    alternative splicing; thyroid peroxidase.
SOURCE      Human thyroid gland (from patients with Grave's disease), cDNA to
            mRNA, clone phTPO-2.8.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 3048)
  AUTHORS   Kimura,S., Kotani,T., McBride,O.W., Umeki,K., Hirai,K., Nakayama,T.
            and Ohtaki,S.
  TITLE     Human thyroid peroxidase: complete cDNA and protein sequence,
            chromosome mapping, and identification of two alternately spliced
            mRNAs
  JOURNAL   Proc. Natl. Acad. Sci. U.S.A. 84 (16), 5555-5559 (1987)
  MEDLINE   87288643
COMMENT     Draft entry and computer-readable sequence for [1] kindly provided
            by S.Kimura, 09-SEP-1987.
FEATURES             Location/Qualifiers
     source          1. .3048
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /map="2pter-p24"
     mRNA            <1. .3048
                     /note="TPO mRNA"
     gene            73. .2874
                     /gene="TPO"
     CDS             73. .2874
                     /gene="TPO"
                     /note="thyroid peroxidase"
                     /codon_start=1
                     /db_xref="GDB:G00-120-446"
                     /protein_id="AAA61215.1"
                     /db_xref="GI:339867"
                     /translation="MRALAVLSVTLVMACTEAFPPPISRGKELLWGKPEESRVSSVLE
                     ESKRLVDTAMYATMQRNLKKRGILSPAQLLSFSKLPEPTSGVIARAAEIMETSIQAMK
                     RKVNLKTQQSQNPTDALSEDLLSIIANMSSCLPYMLPPKCPNTCLANKYRPITGACNN
                     RDBFRWGASNTALARWLPVVYEDGFSQPRGWMGFLYNGPLPVVREVTRHVIQVSNE
```

```
                     VVTDDDRYSDLLMAWGQYIDHDIAFTPQSTSKAAPGGADCQWTCENQNPCFPIQLPE
                     EARPAAGTACLPFTRSSAACQTGDGQALPGNLSTANFRQQMNGLTSFLDASTVYGSSP
                     ALERQLRNWTSASLLRVHARLRDSGRAYLPYVPPRRPAACAPEPGIPGETRGPCFLA
                     GDGRASEVPSLTALHTLWLREHNRLAAALKALNAHWSADAVYQEARKVVGALHQIITL
                     RDYIPRILGPEAFQQYVGPYEGYDSTANPTVSNVFSTAAFRFGHATIHPLVRRLDASF
                     QDHPDLPGLWLHQAFFSPWTLLRGGGLDPLIRGLLARPAKLQVDQLIMNEELTERLFV
                     LSNSSTLDLASINLQRGRDHGLPGYNEWREFCGLPRLETPADLSTAIASRSVADKILD
                     LYKHPDNIDVWLGGLAENFLPRARTGPLFACLIGKQMKALRDGDWFWWENSHVFTDAQ
                     RRELEKHSLSRVICDNTGLTVPVMDAPQVGKFPEDFESCDSIPGMNLEAWRETPPDDD
                     KCGFPESVENGDFVHCEESGRRVLVYSCRHGYELQGREQLTCTQDGWDFPPPLCKDVN
                     ECADGAHPCHASARCRNTKGGFQCLCADPYELGDDGRTCVDSGRLPRATWISMSLAA
                     LLIGGFAGLTSTVICRWFKRSTLPISETGGGTPELRCGRHQAVGTSPQRAAAQDS
                     EQESAGMEGRDTHRLPRAL"
BASE COUNT      671 a    930 c    881 g    566 t
ORIGIN      Chromosome 2pter-qll.


alignment_scores:
                Quality: 4528.00              Length:    848
                  Ratio:  5.365                Gaps:      0
 Percent Similarity:  99.528    Percent Identity:  99.175


alignment_block:
 US-08-482-402A-3_COPY_1_848 x HUMTPOA  ..

 Align seq 1/1  to: HUMTPOA  from: 1  to: 3048

       1 MetArgAlaLeuAlaValLeuSerValThrLeuValMetAlaCysThrGl 17
         ||||||||||||||||||||||||||||||||||||||||||||||||||
      73 ATGAGAGCGCTCGCTGTGGCTGTCTGTCACGCTGGTTATGGCCTGCACAGA 122

      17 uAlaPhePheProPheIleSerArgGlyLysGluLeuLeuTrpGlyLysP 34
         ||||||||||||||||||||||||||||||||||||||||||||||||||
     123 AGCCTTCTTCCCCTTCATCTCGAGAGGGAAAGAACTCCTTTGGGGAAAGC 172

      34 roGluGluSerArgValSerSerValLeuGluGluSerLysArgLeuVal 50
         ||||||||||||||||||||||||||||||||||||||||||||||||||
     173 CTGAGGAGTCTCGTGTCTCTAGCGTCTTGGAGGAAAGCAAGCGCCTGGTG 222

      51 AspThrAlaMetTyrAlaThrMetGlnArgAsnLeuLysLysArgGlyIl 67
         ||||||||||||||||||||||||||||||||||||||||||||||||||
     223 GACACCGCCATGTACGCCACGATGCAGAGAAACCTCAAGAAAAGAGGAAT 272

      67 eLeuSerGlyAlaGlnLeuLeuSerPheSerLysLeuProGluProThrS 84
         |||||||  ||||||||||||||||||||||||||||||||||||||||
     273 CCTTCTCCAGCTCAGCTTCTGTCTTTTCCAACTTCCTCGAGCCAACAA 322

      84 erGlyValIleAlaArgAlaAlaGluIleMetGluThrSerIleGlnAla 100
         ||||||||||||||||||||||||||||||||||||||||||||||||||
     323 GCGGAGTGATTGCCCGAGCAGCAGAGATAATGGAAACATCAATACAAGCG 372

     101 MetLysArgLysValAsnLeuLysThrGlnGlnSerGlnHisProThrAs 117
         ||||||||||||||||||||||||||||||||||||||||||||||||||
     373 ATGAAAAGAAAAGTCAACCTGAAAACTCAACAATCACAGCATCCAACGGA 422

     117 pAlaLeuSerGluAspLeuLeuSerIleIleAlaAsnMetSerGlyCysL 134
         ||||||||||||||||||||||||||||||||||||||||||||||||||
     423 TGCTTTATCAGAAGACTCTGCTGAGCATCATTGCAAACATGTCTGGATGTC 472

     134 euProTyrMetLeuProProLysCysProAsnThrCysLeuAlaAsnLys 150
         ||||||||||||||||||||||||||||||||||||||||||||||||||
     473 TCCCTTACATGCTGCCCCCAAAATGCCCAACACTTGCCTGGCGAACAAA 522

     151 TyrArgProIleThrGlyAlaCysAsnAsnArgAspHisProArgTrpGl 167
         ||||||||||||||||||||||||||||||||||||||||||||||||||
     523 TACAGGCCCATCACAGGAGCTGCAACAACAGAGACCACCCCAGATGGGG 572

     167 yAlaSerAsnThrAlaLeuAlaArgTrpLeuProProValTyrGluAspG 184
         ||||||||||||||||||||||||||||||||||||||||||||||||||
     573 CGCCTCCAACACGGCCCTGGCACGATGGCTCCCTCCAGTCTATGAGGACG 622

     184 lyPheSerGlnProArgGlyTrpAsnProGlyPheLeuTyrAsnGlyPhe 200
         ||||||||||||||||||||||||||||||||||||||||||||||||||
```


623 GCTTCAGTCAGCCCCGAGGCTGGAACCCCGGCTTCTTGTACAACGGGTTC 672
201 ProLeuProProValArgGluValThrArgHisValIleGlnValSerAs 217
673 CCACTGCCCCCGGTCCGGGAGGTGACAAGACATGTCATTCAAGTTTCAAA 722
217 nGluValValThrAspAspAspArgTyrSerAspLeuLeuMetAlaTrpG 234
723 TGAGGTTGTCACAGATGATGACCGCTATTCTGACCTCCTGATGGCATGGG 772
234 lyGlnTyrIleAspHisAspIleAlaPheThrProGlnSerThrSerLys 250
773 GACAATACATCGACCACGACATCGCGTTCACACCACAGAGCACCAGCAAA 822
251 AlaAlaPheGlyGlyGlySerAspCysGlnMetThrCysGluAsnGlnAs 267
823 GCTGCCTTCGGGGGAGGGGCTGACTGCCAGATGACTTGTGAGAACCAAAA 872
267 nProCysPheProIleGlnLeuProGluGluAlaArgProAlaAlaGlyT 284
873 CCCATGTTTTCCCATACAACTCCCGGAGGAGGCCCGGCCGGCGGGCA 922
284 hrAlaCysLeuProPheTyrArgSerSerAlaAlaCysGlyThrGlyAsp 300
923 CCGCCTGTCTGCCCTTCTACCGCTCTTCGGCCGCCTGCGGCACCGGGAC 972
301 GlnGlyAlaLeuPheGlyAsnLeuSerThrAlaAsnProArgGlnGlnMe 317
973 CAAGGCGCGCTCTTTGGGAACCTGTCCACGGCCAACCCGCGGCAGCAGAT 1022
317 tAsnGlyLeuThrSerPheLeuAspAlaSerThrValTyrGlySerSerP 334
1023 GAACGGTTGACCTCGTTCCTGGACGCCTCCACCGTGTACGGCAGCCCC 1072
334 roAlaLeuGluArgGlnLeuArgAsnTrpThrSerAlaGluGlyLeuLeu 350
1073 CGGCCCTAGAGAGGCAGCTGCGGAACTGGACCAGTGCCGAAGGGCTGCTC 1122
351 ArgValHisGlyArgLeuArgAspSerGlyArgAlaTyrLeuProPheVa 367
1123 CGCGTCCACGCGCGCCTCCGGGACTCCGGCCGCCTACCTGCCCTTCGT 1172
367 lProProArgAlaProAlaAlaCysAlaProGluProGlyAsnProGlyG 384
1173 GCCGCCACGGCGCCTKCGGCCTGTGCGCCGACCCGACCCTCCCGGAG 1222
384 luThrArgGlyProCysPheLeuAlaGlyAspGlyArgAlaSerGluVal 400
1223 AGACCCGCGGGCCTGCTTCCTGGCCGGAGACGCCGCCGCCAGCGAGGTC 1272
401 ProSerLeuThrAlaLeuHisThrLeuTrpLeuArgGluHisAsnArgLe 417
1273 CCCTCCCTGACGGCACTGCACACGCTGTGGCTGCGCGAGCACAACCGCCT 1322
417 uAlaAlaAlaLeuLysAlaLeuAsnAlaHisTrpSerAlaAspAlaVal7 434
1323 GGCCGCGGCGCTAAGCGCCTCAATGCGCACTGGAGCCCGGACGCCGTGT 1372
434 yrGlnGluAlaArgLysValValGlyAlaLeuHisGlnIleTleThrLeu 450
1373 ACCAGGAGGCGCGCAAGGTCGTGGGCGCTCTGCACCAGATCATCACCCTG 1422
451 ArgAspTyrTleProArgIleLeuGlyProGluAlaPheGlnGlnTyrVa 467
1423 AGGGATTACATCCCCAGGATCCTGGGACCCGAGGCCTTCCAGCAGTACGT 1472
467 lGlyProTyrGluGlyTyrAspSerThrAlaAsnProThrValSerAsnV 484
1473 GGGTCCCTATGAAGGCTATGACTCCACCGCCAACCCCACTGTGTCCAACG 1522
484 alPheSerThrAlaAlaPheArgPheGlyHisAlaThrIleHisProLeu 500
1523 TGTTCTCCACAGCCGCCTTCCGCTTCGGCCATGCCACGATCCACCCGCTG 1572
501 ValArgArgLeuAspAlaSerPheGlnGluHisProAspLeuProGlyLe 517
1573 GTGAGGAGGCTGGACGCCAGCTTCCAGGAGCACCCGACCTGCCCGGGCT 1622
517 uTrpLeuHisGlnAlaPhePheSerProTrpThrLeuLeuArgGlyGlyG 534
1623 GTGGCTGCACCAGGCTTTCTTCAGCCCATGGACATTACTCCGTGGAGGTG 1672
534 lyLeuAspProLeuIleArgGlyLeuLeuAlaArgProAlaLysLeuGln 550
1673 GTTTGGACCCACTAATACGAGGCCTTCTTCAAGACCAGCCAAACTGCAG 1722
551 ValGlnAspGlnLeuMetAsnGluGluLeuThrGluArgLeuPheValLe 567
1723 GTGCAGGATCAGCTGATGAACGAGGAGCTGACGGAAAGGCTCTTTGTGCT 1772
567 uSerAsnSerSerThrLeuAspLeuAlaSerIleAsnLeuGlnArgGlyA 584
1773 GTCCAATTCCAGCACCTTGGATCTGGCGTCCATCAACCTGCAGAGGGGCC 1822
584 rgAspHisGlyLeuProGlyTyrAsnGluTrpArgGluPheCysGlyLeu 600
1823 GGGACCACGGGCTGCCAGGTTACAATGATGGAGGAGTTCTGCGGCCTG 1872
601 ProArgLeuGluThrProAlaAspLeuSerThrAlaIleAlaSerArgSe 617
1873 CCTCGCCTGGAGACCCCCGCTGACCTGAGCACAGCCATCGCCAGCAGGAG 1922
617 rValAlaAspLysIleLeuAspLeuTyrLysHisProAspAsnIleAspV 634
1923 CGTGGCCGACAAGATCCTGGACTTGTACAAGCATCCTGACAACATCGATG 1972
634 alTrpLeuGlyGlyLeuAlaGlnAsnPheLeuProArgAlaArgThrGly 650
1973 TCTGGCTGGGAGGCTTAGCTGAAAACTTCCTCCCCAGGGCTCGGACAGG 2022
651 ProLeuPheAlaCysLeuIleGlyLysGlnMetLysAlaLeuArgAspGl 667
2023 CCCCTGTTTGCCTGTCTCATTGGGAAGCAGATGAAGGCTCTGCGGGATGG 2072
667 yAspTrpPheTrpTrpGluAsnSerHisValPheThrAspAlaGlnArgA 684
2073 TGACTGGTTTTGGTGGGAGAACAGCCACGTCTTCACGGATGCACAGAGGC 2122
684 rgGluLeuGluLysHisSerLeuSerArgValIleCysAspAsnThrGly 700
2123 GTGAGCTGGAGAAGCACTCCCTGTCTCGGGTCATCTGTGACAACACTGGC 2172
701 LeuThrArgValProMetAspAlaPheGlnValGlyLysPheProGluAs 717
2173 CTCACCAGGGTGCCCATGGATGCCTTCCAAGTCGGCAAATTCCCTGAAGA 2222
717 pPheGluSerCysAspSerIleThrGlyMetAsnLeuGluAlaTrpArgG 734
2223 CTTTGAGTCTTGTGACAGCATCCCTGGCATGAACCTGGAGGCCTGGAGGG 2272
734 luThrPheProGlnAspAspLysCysGlyPhePheProGluSerValGluAsn 750
2273 AAACCTTTCCTCAAGACGACAAGTGTGGCTTCTTCCCAGAGAGCGTGGAGAAT 2322
751 GlyAspPheValHisCysGluGluSerGlyArgArgValLeuValTyrSe 767
2323 GGGGACTTTGTCACTGTGAGGAGTCTGGGAGGCGCGTGCTGGTTATTC 2372
767 rCysArgHisGlyTyrGluLeuGlnGlyArgGluGlnLeuThrCysThrG 784
2373 CTGCCGGCACGGGTATGAGCTCCAAGGCCGGGAGCAGCTCACTTGCACCC 2422
784 lnGluGlyTrpAspPheGlnProProLeuCysLysAspValAsnGluCys 800
2423 AGGAAGGATGGGATTTCCAGCCTCCCCTCTGCAAAGATGTGAACGAGTGT 2472

```
     801 AlaAspGlyAlaHisProProCysHisAlaSerAlaArgCysArgAsnTh 817
         |||||||||||||||||||||||||||||||||||||||||||||||||
    2473 GCAGACGGTGCCCACCCCCCCTGCCAGCGCTCTGCGAGGTGCAGAAACAC 2522

     817 rLysGlyGlyPheGlnCysLeuCysAlaAspProTyrGluLeuGlyAspA 834
         |||||||||||||||||||||||||||||||||||||||||||||||||
    2523 CAAAGGCGGCTTCCAGTGTCTCTGCGCGGACCCCTACGAGTTAGGAGACG 2572

     834 spGlyArgThrCysValAspSerGlyArgLeuProArgValThr 848
         ||||||||||||||||||||||||||||||:::|||
    2573 ATGGGAGAACCTGCGTAGACTCCGGGAGGCTCCCTCGGGCGACT 2616

seq_name: gb_pat:A92132

seq_documentation_block:
LOCUS       A92132          2847 bp    DNA   circular  PAT       22-JAN-2000
DEFINITION  Sequence 1 from Patent WO9820354.
ACCESSION   A92132
VERSION     A92132.1  GI:6740928
KEYWORDS    .
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 2847)
  AUTHORS   Furmaniak,J. and Grennan,J.F.
  TITLE     MONITORING REACTIVITY OF THYROID PEROXIDASE AUTOANTIBODIES
  JOURNAL   Patent: WO 9820354-A 1 14-MAY-1998;
            FURMANIAK JADWIGA (GB); GRENNAN JONES FIONA (GB)
FEATURES             Location/Qualifiers
     source          1. .2847
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
BASE COUNT      610 a    884 c    833 g    520 t
ORIGIN

alignment_scores:
                Quality: 4525.00          Length:    848
                  Ratio:  5.355             Gaps:      0
 Percent Similarity:  99.646    Percent Identity: 99.057

alignment_block:
 US-08-482-402A-3_COPY_1_848 x A92132   ..

 Align seg 1/1  to: A92132  from: 1  to: 2847

       1 MetArgAlaLeuAlaValLeuSerValThrLeuValMetAlaCysThrGl 17
         ||||||||||||||||||||||||||||||||||||||||||||||||||
      46 ATGAGACGGCTCGCTGTGCTGTCTGTCACGCTGGTTATGGCCTGCACAGA 95

      17 uAlaPhePheProPheIleSerArgGlyLysGluLeuLeuTrpGlyLysP 34
         ||||||||||||||||||||||||||||||||||||||||||||||||||
      96 AGCCTTCTCCCCTTCATCTCGAGAGGGAAAGAACTCCTTTGGGGAAAGC 145

      34 roGluGluSerArgValSerSerValLeuGluGluSerLysArgLeuVal 50
         ||||||||||||||||||||||||||||||||||||||||||||||||||
     146 CTGAGGAGTCTCGTCGTCTCTACCGTCTTGAGGAAGCAAGCGCCTGGTG 195

      51 AspThrAlaMetTyrAlaThrMetGlnArgAsnLeuLysLysArgGlyIl 67
         ||||||||||||||||||||||||||||||||||||||||||||||||||
     196 GACACCGCCATGTACGCCACGATGCAGAAGAAACCTCAAGAAAAGAGGAAT 245

      67 eLeuSerGlyAlaGlnLeuLeuSerPheSerLysLeuProGluProThrS 84
         |||||||  ||||||||||||||||||||||||||||||||||||||||
     246 CCTTTCTCCAGCTCAGCTTCTGTCTTTTTCCAACTTCCTGAGCCAACAA 295

      84 erGlyValIleAlaArgAlaAlaGluIleMetGluThrSerIleGlnAla 100
         ||||||||||||||||||||||||||||||||||||||||||||||||||
     296 GCGGAGTGATTGCCCGAGCAGCAGAGATAATGGAAACATCAATACAAGCG 345

     101 MetLysArgLysValAsnLeuLysThrGlnGlnSerGlnHisProThrAs 117
         ||||||||||||||||||||||||||||||||||||||||||||||||||

     346 ATGAAAAGAAAAGTCAACCTGAAAACTCAACAATCACAGCATCCAACGGA 395

     117 pAlaLeuSerGluAspLeuLeuSerIleIleAlaAsnMetSerGlyCysL 134
         ||||||||||||||||||||||||||||||||||||||||||||||||||
     396 TGCTTTATCAGAAGATCTGCTGAGCATCATTGCAAACATGTCTGGATGTC 445

     134 euProTyrMetLeuProProLysCysProAsnThrCysLeuAlaAsnLys 150
         ||||||||||||||||||||||||||||||||||||||||||||||||||
     446 TCCCTTACATGCTGCCCCAAAATGCCCAACACTTGCCTGCCGAACAAA 495

     151 TyrArgProIleThrGlyAlaCysAsnAsnArgAspHisProArgTrpGl 167
         ||||||||||||||||||||||||||||||||||||||||||||||||||
     496 TACAGGCCCATCACAGGAGCTTGCAACAACAGAGACCACCCCAGATGGGG 545

     167 yAlaSerAsnThrAlaLeuAlaArgTrpLeuProProValTyrGluAspG 184
         ||||||||||||||||||||||||||||||||||||||||||||||||||
     546 CGCCTCCAACACGGCCCTGGCACGATGGCTCCCTCCAGTCTATGAGGACG 595

     184 lyPheSerGlnProArgGlyTrpAsnProGlyPheLeuTyrAsnGlyPhe 200
         ||||||||||||||||||||||||||||||||||||||||||||||||||
     596 GCTTCAGTCAGCCCGGAGGCTGGAACCCGGCTTCTTGTACAACGGGTTC 645

     201 ProLeuProProValArgGluValThrArgHisValIleGlnValSerAs 217
         ||||||||||||||||||||||||||||||||||||||||||||||||||
     646 CCACTGCCCCCGGTCCGGGAGGTGACAAGACATGTCATTCAAGTTTCAAA 695

     217 nGluValValThrAspAspAspArgTyrSerAspLeuLeuMetAlaTrpG 234
         ||||||||||||||||||||||||||||||||||||||||||||||||||
     696 TGAGGTTGTCACAGATGATGACCGCTATTCTGACCTCCTGATGGCATGGG 745

     234 lyGlnTyrIleAspHisAspIleAlaPheThrProGlnSerThrSerLys 250
         ||||||||||||||||||||||||||||||||||||||||||||||||||
     746 GACAATACATCGACCACGACATCGCGTTCACACCACAGAGCACCAGCAAG 795

     251 AlaAlaPheGlyGlyGlySerAspCysGlnMetThrCysGluAsnGlnAs 267
         ||||||||||||||||||:::|||||||||||||||||||||||||||||
     796 GCTCCCTCGGGGGAGGGGCTGACTGCCAGATACTTCTGTGAGACCAAAA 845

     267 nProCysPheProIleGlnLeuProGluGluAlaArgProAlaAlaGlyT 284
         ||||||||||||||||||||||||||||||||||||||||||||||||||
     846 CCCATGTTTTCCCATACAACTCCCGGAGGAGGCCCGGCCGGCCGGGCA 895

     284 hrAlaCysLeuProPheTyrArgSerSerAlaAlaCysGlyThrGlyAsp 300
         ||||||||||||||||||||||||||||||||||||||||||||||||||
     896 CCGCCTGTCGCCCTTCTACCGCTCTTCGGCCGCTGCGGCACCGGGGAC 945

     301 GlnGlyAlaLeuPheGlyAsnLeuSerThrAlaAsnProArgGlnGlnMe 317
         ||||||||||||||||||||||||||||||||||||||||||||||||||
     946 CAAGGCGGCCTCTTTGGGAACCTGTCCACGGCCAACCCGCGGCAGCAGAT 995

     317 tAsnGlyLeuThrSerPheLeuAspAlaSerThrValTyrGlySerSerP 334
         ||||||||||||||||||||||||||||||||||||||||||||||||||
     996 GAACGGGTTGACCTCGTTCCTGGACGCGTCCACCGTGTATGGCAGCTCCC 1045

     334 roAlaLeuGluArgGlnLeuArgAsnTrpThrSerAlaGluGlyLeuLeu 350
         ||||||||||||||||||||||||||||||||||||||||||||||||||
    1046 CGGCCCTAGAGAGGCAGCTGCGGAACTGGACCAGTGCGAAGGGCTGCTC 1095

     351 ArgValHisGlyArgLeuArgAspSerGlyArgAlaTyrLeuProPheVa 367
         ||||||||:::|||||||||||||||||||||||||||||||||||||||
    1096 GGCGTCACGCCGCCTCCGGGACTCCGGCCGCCGCCTACCTGCCCTTCGT 1145

     367 lProProArgAlaProAlaAlaCysAlaProGlyProGlyAsnProGlyG 384
         ||||||||||||||||||||||||||||||||||||||||||||||||||
    1146 GCCGCCACGCGCGCCTGCGGCCTGTGCGCCCGAGCCCGGCATCCCCGGAG 1195

     384 luThrArgGlyProCysPheLeuAlaGlyAspGlyArgAlaSerGluVal 400
         ||||||||||||||||||||||||||||||||||||||||||||||||||
    1196 AGACCCGCGGGCCCTGCTTCCTGGCCGGAGACGGCCGCCGCCAGCGAGGTC 1245

     401 ProSerLeuThrAlaLeuHisThrLeuTrpLeuArgGluHisAsnArgLe 417
         ||||||||||||||||||||||||||||||||||||||||||||||||||
    1246 CCCTCCCTGACGGCACTGCACACGCTGTGGCTGCGCGAGCACAACCGCCT 1295
```




```
seq_name: gb_pr:HSTPO

seq_documentation_block:
LOCUS       HSTPO      3027 bp    mRNA            PRI       23-MAR-1995
DEFINITION  Human mRNA for thyroperoxidase.
ACCESSION   Y00406
VERSION     Y00406.1  GI:37250
KEYWORDS    membrane glycoprotein; thyroperoxidase.
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 3027)
  AUTHORS   Libert,F., Ruel,J., Ludgate,M., Swillens,S., Alexander,N.,
            Vassart,G. and Dinsart,C.
  TITLE     Complete nucleotide sequence of the human
            thyroperoxidase-microsomal antigen cDNA
  JOURNAL   Nucleic Acids Res. 15 (16), 6735 (1987)
  MEDLINE   87316933
REFERENCE   2  (bases 1 to 3027)
  AUTHORS   Libert,F.
  TITLE     Direct Submission
  JOURNAL   Submitted (28-OCT-1987) to the EMBL/GenBank/DDBJ databases
FEATURES             Location/Qualifiers
     source          1..3027
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /tissue_type="thyroid"
                     /clone_lib="lambda gt11"
                     /map="chromosome 2, short arm"
                     /clone="L 2842
     CDS             /note="precursor polypeptide"
                     /codon_start=1
                     /protein_id="CAA58467.1"
                     /db_xref="GI:37251"
                     /db_xref="SWISS-PROT:P07202"
                     /translation="MRALAVLSVTLVMACTEAFFPFISRGKELIWGKPEESRVSVLF
                     ESKRLVDTAMYATMQRNLKKRSILSPAQLLSFSKLPEPTSGVIARAAEIMETSIQAMK
                     RKVNLKTQQSQMPTDALSEDLLSIIANMSGCLPYMLPPKCPNTCLANKYRPITGACNN
                     RDHPRWGASNTALARWLPPVYEDGFSQPRGWNPGFLYNGFPLPPVREVTRHVIQVSNE
                     VVTDDRRYSDLLMAWGQYIDHDIAFTPQSTSKAAFGGGADCQMFCENQNPCFPIQLPE
```

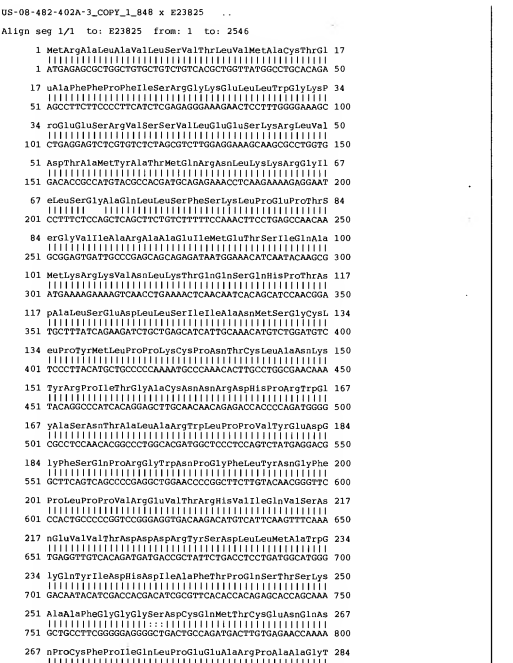
```
                   EARPAAGTACLPFYRSSAACGTGDQGALFGNLSTANPRQQMNGLTSFLDASTVYGSSP
                   ALERQLRNWTSAEGLLRVHARLRDSGRAYLPFVPPRAPAACAPEPGIPGETRGPCFLA
                   GDGRASEVPSLTALHTLWLREHNRLAAALKALNAHWSADAVYQEARKVVGALHQIITL
                   RDYIPRILGPEAFQQYVGPYEGYDSTANPTVSNVFSTAAFRFGHATIHPLVRRLDASF
                   QEHPDLPGLWLHQAFFSPWTLLRGGGLDPLIRGLLARPAKLQVQDQLMNEELTERLFV
                   LSNSSTLNLASINLQRGRDHGLPGYNEWREFCGLPRLETPADLSTAIASRSVADKILD
                   LYKHPDNIDVWLGGLAENFLPRARTGPLFACLIGKQMKALRDGDWFWWENSHVFTDAQ
                   RRELEKHSLSRVICDNTGLTRVPMDAFQVGKFPEDFESCDSIPGMNLEAWRETFPQDD
                   KCGFPESMENGDFVHCEESGRRVLVYSCRHGYELQGREQLTCTQEGWDFQPPLCKDVN
                   ECADGAHPPCHASARCRNTKGGFQCLCADPYELGDDGRTCVDSGRLPRATWISMSLAA
                   LLIGGFAGLTSTVICKWTRTGTKSTLPISETGGGTPELRCGKHQAVGTSPQRAAAQDS
                   EQESAGMEGRDTHRLPRAL"
     sig_peptide   41. .82
                   /product="put. signal peptide (AA -14 to -1)"
     mat_peptide   83. .2839
                   /product="put. mature thyroperoxidase (AA 1-919)"
     misc_feature  425. .433
                   /note="pot. N-glycosylation site"
     misc_feature  959. .967
                   /note="pot. N-glycosylation site"
     misc_feature  1064. .1072
                   /note="pot. N-glycosylation site"
     misc_feature  1472. .1480
                   /note="pot. N-glycosylation site"
     misc_feature  1745. .1753
                   /note="pot. N-glycosylation site"
     misc_feature  2579. .2652
                   /note="pot. transmembrane region"
BASE COUNT      664 a    928 c    871 g    564 t
ORIGIN

alignment_scores:
          Quality: 4525.00            Length:     848
            Ratio:  5.355               Gaps:       0
 Percent Similarity:  99.646   Percent Identity: 99.057

alignment_block:
 US-08-482-402A-3_COPY_1_848 x HSTPO   ..

 Align seg 1/1  to: HSTPO  from: 1  to: 3027

       1 MetArgAlaLeuAlaValLeuSerValThrLeuValMetAlaCysThrGl 17
         ||||||||||||||||||||||||||||||||||||||||||||||||||
      41 ATGAGAGCGCTCGCTGTGCTGTCTGTCACGCTGGTTATGGCCTGCACAGA 90

      17 uAlaPhePheProPheIleSerArgGlyLysGluLeuLeuTrpGlyLysP 34
         ||||||||||||||||||||||||||||||||||||||||||||||||||
      91 AGCCTTCTTCCCCTTCATCTCGAGAGGGAAAGAACTCCTTTGGGGAAAGC 140

      34 roGluGluSerArgValSerSerValLeuGluGluSerLysArgLeuVal 50
         ||||||||||||||||||||||||||||||||||||||||||||||||||
     141 CTGAGGAGTCTCGTGTCTCTAGCGTCTTGGAGGAAAGCAAGCGCCTGGTG 190

      51 AspThrAlaMetTyrAlaThrMetGlnArgAsnLeuLysLysArgGlyIl 67
         ||||||||||||||||||||||||||||||||||||||||||||||||||
     191 GACACCGCCATGTACGCCACGATGCAGAGAAACCTCAAGAAAAGAGGAAT 240

      67 eLeuSerGlyAlaGlnLeuLeuSerPheSerLysLeuProGluProThrS 84
         ||||||    ||||||||||||||||||||||||||||||||||||||||
     241 CCTTTCTCCAGCTCAGCTTCTGTCTTTTTCCAAACTTCCTGAGCCAACAA 290

      84 erGlyValIleAlaArgAlaAlaGluIleMetGluThrSerIleGlnAla 100
         ||||||||||||||||||||||||||||||||||||||||||||||||||
     291 GCGGAGTGATTGCCCGAGCAGCAGAGATAATGGAAACATCAATACAAGCG 340

     101 MetLysArgLysValAsnLeuLysThrGlnGlnSerGlnHisProThrAs 117
         ||||||||||||||||||||||||||||||||||||||||||||||||||
     341 ATGAAAAGAAAAGTCAACCTGAAAACTCAACAATCACAGCATCCAACGGA 390

     117 pAlaLeuSerGluAspLeuLeuSerIleIleAlaAsnMetSerGlyCysL 134
         ||||||||||||||||||||||||||||||||||||||||||||||||||
     391 TGCTTTATCAGAAGATCTGCTGAGCATCATTGCAAACATGTCTGGATGTC 440

     134 euProTyrMetLeuProProLysCysProAsnThrCysLeuAlaAsnLys 150
         ||||||||||||||||||||||||||||||||||||||||||||||||||
     441 TCCCTTACATGCTGCCCCCAAAATGCCCAAACACTTGCCTGGCGAACAAA 490

     151 TyrArgProIleThrGlyAlaCysAsnAsnArgAspHisProArgTrpGl 167
         ||||||||||||||||||||||||||||||||||||||||||||||||||
     491 TACAGGCCCATCACAGGAGCTTGCAACAACAGAGACCACCCCAGATGGGG 540

     167 yAlaSerAsnThrAlaLeuAlaArgTrpLeuProProValTyrGluAspG 184
         ||||||||||||||||||||||||||||||||||||||||||||||||||
     541 CGCCTCCAACACGGCCCTGGCACGATGGCTCCCTCCAGTCTATGAGGACG 590

     184 lyPheSerGlnProArgGlyTrpAsnProGlyPheLeuTyrAsnGlyPhe 200
         ||||||||||||||||||||||||||||||||||||||||||||||||||
     591 GCTTCAGTCAGCCCCGAGGCTGGAACCCCGGCTTCTTGTACAACGGGTTC 640

     201 ProLeuProProValArgGluValThrArgHisValIleGlnValSerAs 217
         ||||||||||||||||||||||||||||||||||||||||||||||||||
     641 CCACTGCCCCCGGTCCGGGAGGTGACAAGACATGTCATTCAAGTTTCAAA 690

     217 nGluValValThrAspAspAspArgTyrSerAspLeuLeuMetAlaTrpG 234
         ||||||||||||||||||||||||||||||||||||||||||||||||||
     691 TGAGGTTGTCACAGATGATGACCGCTATTCTGACCTCCTGATGGCATGGG 740

     234 lyGlnTyrIleAspHisAspIleAlaPheThrProGlnSerThrSerLys 250
         ||||||||||||||||||||||||||||||||||||||||||||||||||
     741 GACAATACATCGACCACGACATCGCGTTCACACCACAGAGCACCAGCAAA 790

     251 AlaAlaPheGlyGlyGlySerAspCysGlnMetThrCysGluAsnGlnAs 267
         |||||||||||||||||:::||||||||||||||||||||||||||||||
     791 GCTGCCTTCGGGGGAGGGGCTGACTGCCAGATGACTTGTGAGAACCAAAA 840

     267 nProCysPheProIleGlnLeuProGluGluAlaArgProAlaAlaGlyT 284
         ||||||||||||||||||||||||||||||||||||||||||||||||||
     841 CCCATGTTTTCCCATACAACTCCCGGAGGAGGCCCGGCCGGCCGCGGGCA 890

     284 hrAlaCysLeuProPheTyrArgSerSerAlaAlaCysGlyThrGlyAsp 300
         ||||||||||||||||||||||||||||||||||||||||||||||||||
     891 CCGCCTGTCTGCCCTTCTACCGCTCTTCGGCCGCCTGCGGCACCGGGGAC 940

     301 GlnGlyAlaLeuPheGlyAsnLeuSerThrAlaAsnProArgGlnGlnMe 317
         ||||||||||||||||||||||||||||||||||||||||||||||||||
     941 CAAGGCGCGCTCTTTGGGAACCTGTCCACGGCCAACCCGCGGCAGCAGAT 990

     317 tAsnGlyLeuThrSerPheLeuAspAlaSerThrValTyrGlySerSerP 334
         ||||||||||||||||||||||||||||||||||||||||||||||||||
     991 GAACGGGTTGACCTCGTTCCTGGACGCGTCCACCGTGTATGGCAGCTCCC 1040

     334 roAlaLeuGluArgGlnLeuArgAsnTrpThrSerAlaGluGlyLeuLeu 350
         ||||||||||||||||||||||||||||||||||||||||||||||||||
    1041 CGGCCCTAGAGAGGCAGCTGCGGAACTGGACCAGTGCCGAAGGGCTGCTC 1090

     351 ArgValHisGlyArgLeuArgAspSerGlyArgAlaTyrLeuProPheVa 367
         ||||||||:::|||||||||||||||||||||||||||||||||||||||
    1091 CGCGTCCACGCGCGCCTCCGGGACTCCGGCCGCGCCTACCTGCCCTTCGT 1140

     367 lProProArgAlaProAlaAlaCysAlaProGluProGlyAsnProGlyG 384
         ||||||||||||||||||||||||||||||||||||||   |||||||||
    1141 GCCGCCACGCGCGCCTGCGGCCTGTGCGCCCGAGCCCGGCATCCCCGGAG 1190

     384 luThrArgGlyProCysPheLeuAlaGlyAspGlyArgAlaSerGluVal 400
         ||||||||||||||||||||||||||||||||||||||||||||||||||
    1191 AGACCCGCGGGCCCTGCTTCCTGGCCGGAGACGGCCGCGCCAGCGAGGTC 1240

     401 ProSerLeuThrAlaLeuHisThrLeuTrpLeuArgGluHisAsnArgLe 417
         ||||||||||||||||||||||||||||||||||||||||||||||||||
    1241 CCCTCCCTGACGGCACTGCACACGCTGTGGCTGCGCGAGCACAACCGCCT 1290

     417 uAlaAlaAlaLeuLysAlaLeuAsnAlaHisTrpSerAlaAspAlaValT 434
         ||||||||||||||||||||||||||||||||||||||||||||||||||
    1291 GGCCGCGGCGCTCAAGGCCCTCAATGCGCACTGGAGCGCGGACGCCGTGT 1340
```

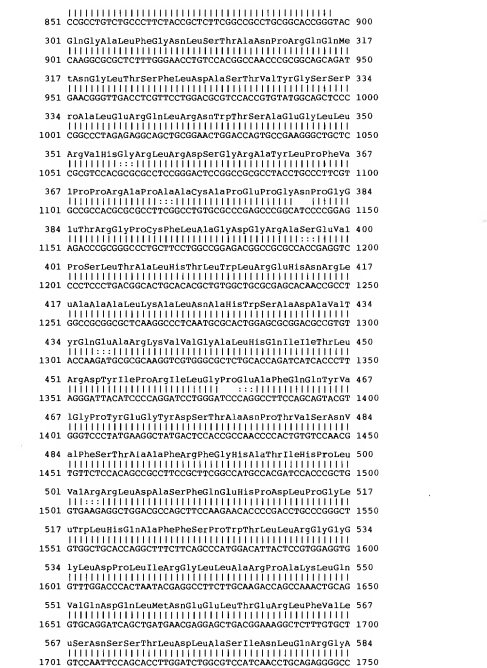
```
 434 yrGlnGluAlaArgLysValValGlyAlaLeuHisGlnIleIleThrLeu 450
     ||||||||||||||||||||||||||||||||||||||||||||||||||
1341 ACCAGGAGGCGCGCAAGGTCGTGGGCGCTCTGCACCAGATCATCACCCTG 1390

 451 ArgAspTyrIleProArgIleLeuGlyProGluAlaPheGlnGlnTyrVa 467
     ||||||||||||||||||||||||||||||||||||||||||||||||||
1391 AGGGATTACATCCCCAGGATCCTGGGACCCGAGGCCTTCCAGCAGTACGT 1440

 467 lGlyProTyrGluGlyTyrAspSerThrAlaAsnProThrValSerAsnV 484
     ||||||||||||||||||||||||||||||||||||||||||||||||||
1441 GGGTCCCTATGAAGGCTATGACTCCACCGCCAACCCCACTGTGTCCAACG 1490

 484 alPheSerThrAlaAlaPheArgPheGlyHisAlaThrIleHisProLeu 500
     ||||||||||||||||||||||||||||||||||||||||||||||||||
1491 TGTTCTCCACAGCCGCCTTCCGCTTCGGCCATGCCACGATCCACCCGCTG 1540

 501 ValArgArgLeuAspAlaSerPheGlnGluHisProAspLeuProGlyLe 517
     ||||||||||||||||||||||||||||||||||||||||||||||||||
1541 GTGAGGAGGCTGGACGCCAGCTTCCAGGAGCACCCCGACCTGCCCGGGCT 1590

 517 uTrpLeuHisGlnAlaPhePheSerProTrpThrLeuLeuArgGlyGlyG 534
     ||||||||||||||||||||||||||||||||||||||||||||||||||
1591 GTGGCTGCACCAGGCTTTCTTCAGCCCATGGACATTACTCCGTGGAGGTG 1640

 534 lyLeuAspProLeuIleArgGlyLeuLeuAlaArgProAlaLysLeuGln 550
     ||||||||||||||||||||||||||||||||||||||||||||||||||
1641 GTTTGGACCCACTAATACGAGGCCTTCTTGCAAGACCAGCCAAACTGCAG 1690

 551 ValGlnAspGlnLeuMetAsnGluGluLeuThrGluArgLeuPheValLe 567
     ||||||||||||||||||||||||||||||||||||||||||||||||||
1691 GTGCAGGATCAGCTGATGAACGAGGAGCTGACGGAAAGGCTCTTTGTGCT 1740

 567 uSerAsnSerSerThrLeuAspLeuAlaSerIleAsnLeuGlnArgGlyA 584
     ||||||||||||||||||:::|||||||||||||||||||||||||||||
1741 GTCCAATTCCAGCACCTTGAATCTGGCGTCCATCAACCTGCAGAGGGGCC 1790

 584 rgAspHisGlyLeuProGlyTyrAsnGluTrpArgGluPheCysGlyLeu 600
     ||||||||||||||||||||||||||||||||||||||||||||||||||
1791 GGGACCACGGGCTGCCAGGTTACAATGAGTGGAGGGAGTTCTGCGGCCTG 1840

 601 ProArgLeuGluThrProAlaAspLeuSerThrAlaIleAlaSerArgSe 617
     ||||||||||||||||||||||||||||||||||||||||||||||||||
1841 CCTCGCCTGGAGACCCCCGCTGACCTGAGCACAGCCATCGCCAGCAGGAG 1890

 617 rValAlaAspLysIleLeuAspLeuTyrLysHisProAspAsnIleAspV 634
     ||||||||||||||||||||||||||||||||||||||||||||||||||
1891 CGTGGCCGACAAGATCCTGGACTTGTACAAGCATCCTGACAACATCGATG 1940

 634 alTrpLeuGlyGlyLeuAlaGluAsnPheLeuProArgAlaArgThrGly 650
     ||||||||||||||||||||||||||||||||||||||||||||||||||
1941 TCTGGCTGGGAGGCTTAGCTGAAAACTTCCTCCCCAGGGCTCGGACAGGG 1990

 651 ProLeuPheAlaCysLeuIleGlyLysGlnMetLysAlaLeuArgAspGl 667
     ||||||||||||||||||||||||||||||||||||||||||||||||||
1991 CCCCTGTTTGCCTGTCTCATTGGGAAGCAGATGAAGGCTCTGCGGGATGG 2040

 667 yAspTrpPheTrpTrpGluAsnSerHisValPheThrAspAlaGlnArgA 684
     ||||||||||||||||||||||||||||||||||||||||||||||||||
2041 TGACTGGTTTTGGTGGGAGAACAGCCACGTCTTCACGGATGCACAGAGGC 2090

 684 rgGluLeuGluLysHisSerLeuSerArgValIleCysAspAsnThrGly 700
     ||||||||||||||||||||||||||||||||||||||||||||||||||
2091 GTGAGCTGGAGAAGCACTCCCTGTCTCGGGTCATCTGTGACAACACTGGC 2140

 701 LeuThrArgValProMetAspAlaPheGlnValGlyLysPheProGluAs 717
     ||||||||||||||||||||||||||||||||||||||||||||||||||
2141 CTCACCAGGGTGCCCATGGATGCCTTCCAAGTCGGCAAATTCCCTGAAGA 2190

 717 pPheGluSerCysAspSerIleThrGlyMetAsnLeuGluAlaTrpArgG 734
     ||||||||||||||||||||    ||||||||||||||||||||||||||
2191 CTTTGAGTCTTGTGACAGCATCCCTGGCATGAACCTGGAGGCCTGGAGGG 2240

 734 luThrPheProGlnAspAspLysCysGlyPheProGluSerValGluAsn 750
     ||||||||||||||||||||||||||||||||||||||||||||:::||||||
2241 AAACCTTTCCTCAAGACGACAAGTGTGGCTTCCCAGAGAGCATGGAGAAT 2290

 751 GlyAspPheValHisCysGluGluSerGlyArgArgValLeuValTyrSe 767
     ||||||||||||||||||||||||||||||||||||||||||||||||||
2291 GGGGACTTTGTGCACTGTGAGGAGTCTGGGAGGCGCGTGCTGGTGTATTC 2340

 767 rCysArgHisGlyTyrGluLeuGlnGlyArgGluGlnLeuThrCysThrG 784
     ||||||||||||||||||||||||||||||||||||||||||||||||||
2341 CTGCCGGCACGGGTATGAGCTCCAAGGCCGGGAGCAGCTCACTTGCACCC 2390

 784 lnGluGlyTrpAspPheGlnProProLeuCysLysAspValAsnGluCys 800
     ||||||||||||||||||||||||||||||||||||||||||||||||||
2391 AGGAAGGATGGGATTTCCAGCCTCCCCTCTGCAAAGATGTGAACGAGTGT 2440

 801 AlaAspGlyAlaHisProProCysHisAlaSerAlaArgCysArgAsnTh 817
     ||||||||||||||||||||||||||||||||||||||||||||||||||
2441 GCAGACGGTGCCCACCCCCCCTGCCACGCCTCTGCGAGGTGCAGAAACAC 2490

 817 rLysGlyGlyPheGlnCysLeuCysAlaAspProTyrGluLeuGlyAspA 834
     ||||||||||||||||||||||||||||||||||||||||||||||||||
2491 CAAAGGCGGCTTCCAGTGTCTCTGCGCGGACCCCTACGAGTTAGGAGACG 2540

 834 spGlyArgThrCysValAspSerGlyArgLeuProArgValThr 848
     ||||||||||||||||||||||||||||||||||||||:::|||
2541 ATGGGAGAACCTGCGTAGACTCCGGGAGGCTCCCTCGGGCGACT 2584
```

```
seq_name: gb_pat:E23825

seq_documentation_block:
LOCUS       E23825       2546 bp    DNA             PAT       07-FEB-2001
DEFINITION  Antigen for immunoassaying antihuman thyroid peroxidase antibody
            and recombinant human thyroid peroxidase.
ACCESSION   E23825
VERSION     E23825.1  GI:13024570
KEYWORDS    JP 1999094833-A/1.
SOURCE      unidentified.
  ORGANISM  unidentified
            unclassified.
REFERENCE   1  (bases 1 to 2546)
  AUTHORS   Masao,F.
  TITLE     Antigen for immunoassaying antihuman thyroid peroxidase antibody
            and recombinant human thyroid peroxidase
  JOURNAL   Patent: JP 1999094833-A 1 09-APR-1999;
            SRL INC
COMMENT     OS   Unidentified
            PN   JP 1999094833-A/1
            PD   09-APR-1999
            PF   19-SEP-1997 JP 1997273743
            PR
            PI   MASAO FUKUSHIMA
            PC   G01N33/53,C12N9/08,C12N15/09,G01N33/531,G01N33/564//(C12N9/08,
            PC   C12R1:91),
            PC   C12N15/00
            CC   Strandedness: Double;
            CC   Topology: Linear;
            FH   Key             Location/Qualifiers
            FT   source          1. .2546
            FT                   /organism='Unidentified'.
FEATURES             Location/Qualifiers
     source          1. .2546
                     /organism="unidentified"
                     /db_xref="taxon:32644"
BASE COUNT      542 a    804 c    725 g    475 t
ORIGIN


alignment_scores:
            Quality: 4495.00            Length:     848
              Ratio:  5.338               Gaps:       0
 Percent Similarity:  99.292   Percent Identity: 98.349

alignment_block:
```

US-08-482-402A-3_COPY_1_848 x E23825 ..

Align seq 1/1 to: E23825 from: 1 to: 2546


1 MetArgAlaLeuAlaValLeuSerValThrLeuValMetAlaCysThrGl 17
1 ATGAGAGCGCTGGCTGTGCTGTCTGTCACGCTGGTTATGGCCTGCACAGA 50
17 uAlaPhePheProPheIleSerArgGlyLysGluLeuLeuTrpGlyLysP 34
101 AGCCTTCTTCCCCTTCATCTCGAGAGGGAAAGAACTCCTTTGGGGAAAGC 100
34 roGluGluSerArgValSerValLeuGluGluSerLysArgLeuVal 50
101 CTGAGGAGTCTCGTGTCTCAGCGTCTTGGAAGAAAGCAAGCGCCTGGTG 150
51 AspThrAlaMetTyrAlaThrMetGlnArgAsnLeuLysLysArgGlyIl 67
151 GACACCGCCATGTACGCCACGATGCAGAGAAACCTCAAGAAAAGAGGAAT 200
67 eLeuSerGlyAlaGlnLeuLeuSerPheSerLysLeuProGluProThrS 84
201 CCTTTCTCCAGCTCAGCTTCTGTCTTTTCCAAACTTCCTGAGCCAACAA 250
84 erGlyValIleAlaArgAlaAlaGluIleMetGluThrSerIleGlnAla 100
251 GCGGAGTGATTGCCTCGAGCAGCAGAGATAATGGAAACATCAATACAAGG 300
101 MetLysArgLysValAsnLeuLysThrGlnGlnSerGlnHisProThrAs 117
301 ATGAAAAGAAAAGTCAACCTGAAAACTCACAATCACAGCATCCAACGGA 350
117 pAlaLeuSerGluAspLeuLeuSerIleIleAlaAsnMetSerGlyCysL 134
351 TGCTTTATCAGAAGATCTGCTGAGCATCATTGCAAACATGTCTGGATGTC 400
134 euProTyrMetLeuProProLysCysProAsnThrCysLeuAlaAsnLys 150
401 TCCCTTACATGCTGCCCCCAAAATGCCCAAACACTTGCCTGGCGAACAAA 450
151 TyrArgProIleThrGlyAlaCysAsnAsnArgAspHisProArgTrpGl 167
451 TACAGGCCATCACAGGAGCTTGCAACAACAGAGACCACCCAGATGGGG 500
167 yAlaSerAsnThrAlaLeuAlaArgTrpLeuProProValTyrGluAspG 184
501 CGCCTCCAACACGGCCCTGGCACGATGGCTCCCTCCAGTCTATGAGGACG 550
184 lyPheSerGlnProArgGlyTrpAsnProGlyPheLeuTyrAsnGlyPhe 200
551 GCTTCAGTCAGCCCGAGGCTGAACCCCGGCTTCTTGTACAACGGGTTC 600
201 ProLeuProProValArgGluValThrArgHisValIleGlnValSerAs 217
601 CCACTGCCCCCGGTCCGGGAGGTGACAAGACATGTCATTCAAGTTTCAAA 650
217 nGluValValThrAspAspAspArgTyrSerAspLeuLeuMetAlaTrpG 234
651 TGAGGTTGTCACAGATGATGACCGCTATTCTGACCTCCTGATGGCATGGG 700
234 lyGlnTyrIleAspHisAspIleAlaPheThrProGlnSerThrSerLys 250
701 GACAATACATCGACCACGACATCGCGTTCACACCACAGAGCACCAGCAAA 750
251 AlaAlaPheGlyGlyGlySerAspCysGlnMetThrCysGluAsnGlnAs 267
751 GCTGCCTTCGGGGGAGGGGCTGACTGCCAGATGACTTGTGAGAACCAAA 800
267 nProCysPheProIleGlnLeuProGluGluAlaArgProAlaAlaGlyT 284
801 CCCATGTTTTCCCATACAACTCCCGGAGGAGGCCCGGCCGGCCGGGCA 850
284 hrAlaCysLeuProPheTyrArgSerSerAlaAlaCysGlyThrGlyAsp 300

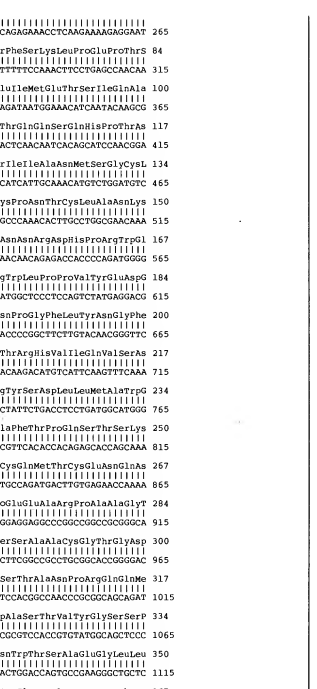


```
1751 GGGACCACGGGCTGCCAGGTTACAATGAGTGGAGGGAGTTCTGCGGCCTG 1800

 601 ProArgLeuGluThrProAlaAspLeuSerThrAlaIleAlaSerArgSe 617
     ||||||||||||||||||||||||||||||||||||||||||||||||||
1801 CCTCGCCTGGAGACCCCCGCTGACCTGAGCACAGCCATCGCCAGCAGGAG 1850

 617 rValAlaAspLysIleLeuAspLeuTyrLysHisProAspAsnIleAspV 634
     ||||||||||||||||||||||||||||||||||||||||||||||||||
1851 CGTGGCCGACAAGATCCTGGACTTGTACAAGCATCCTGACAACATCGATT 1900

 634 alTrpLeuGlyGlyLeuAlaGluAsnPheLeuProArgAlaArgThrGly 650
      |||||||||||||||||||||||||||||||||||||||||||||||||
1901 TCTGGCTGGGAGGCTTAGCTGAAAACTTCCTCCCCAGGGCTCGGACAGGG 1950

 651 ProLeuPheAlaCysLeuIleGlyLysGlnMetLysAlaLeuArgAspGl 667
     ||||||||||||||||||||||||||||||||||||||||||||||||||
1951 CCCCTGTTTGCCTGTCTCATTGGGAAGCAGATGAAGGCTCTGCGGGACGG 2000

 667 yAspTrpPheTrpTrpGluAsnSerHisValPheThrAspAlaGlnArgA 684
     |   ||||||||||||||||||||||||||||||||||||||||||||||
2001 TCACTGGTTTTGGTGGGAGAACAGCCACGTCTTCACGGATGCACAGAGGC 2050

 684 rgGluLeuGluLysHisSerLeuSerArgValIleCysAspAsnThrGly 700
     ||||||||||||||||||||||||||||||||||||||||||||||||||
2051 GTGAGCTGGAGAAGCACTCCCTGTCTCGGGTCATCTGTGACAACACTGGC 2100

 701 LeuThrArgValProMetAspAlaPheGlnValGlyLysPheProGluAs 717
     ||||||||||||||||||||||||||||||||||||||||||||||||||
2101 CTCACCAGGGTGCCCATGGATGCCTTCCAAGTCGGCAAATTCCCCGAAGA 2150

 717 pPheGluSerCysAspSerIleThrGlyMetAsnLeuGluAlaTrpArgG 734
     ||||||||||||||||||||||||||||||||||||||||||||||||||
2151 CTTTGAGTCTTGTGACAGCATCACTGGCATGAACCTGGAGGCCTGGAGGG 2200

 734 luThrPheProGlnAspAspLysCysGlyPheProGluSerValGluAsn 750
     ||||||||||||||||||||||||||||||||||||||||||||||||||
2201 AAACCTTTCCTCAAGACGACAAGTGTGGCTTCCCAGAGAGCGTGGAGAAT 2250

 751 GlyAspPheValHisCysGluGluSerGlyArgArgValLeuValTyrSe 767
     ||||||||||||||||||||||||||||||||||||||||||||||||||
2251 GGGGACTTTGTGCACTGTGAGGAGTCTGGGAGGCGCGTGCTGGTGTATTC 2300

 767 rCysArgHisGlyTyrGluLeuGlnGlyArgGluGlnLeuThrCysThrG 784
     ||||||||||||||||||||||||||||||||||||||||||||||||||
2301 CTGCCGGCACGGGTATGAGCTCCAAGGCCGGGAGCAGCTCACTTGCACCC 2350

 784 lnGluGlyTrpAspPheGlnProProLeuCysLysAspValAsnGluCys 800
     ||||||||||||||||||||||||||||||||||||||||||||||||||
2351 AGGAAGGATGGGATTTCCAGCCTCCCCTCTGCAAAGATGTGAACGAGTGT 2400

 801 AlaAspGlyAlaHisProProCysHisAlaSerAlaArgCysArgAsnTh 817
     ||||||||||||||||||||||||||||||||||||||||||||||||||
2401 GCAGACGGTGCCCACCCCCCCTGCCACGCCTCTGCGAGGTGCAGAAACAC 2450

 817 rLysGlyGlyPheGlnCysLeuCysAlaAspProTyrGluLeuGlyAspA 834
     ||||||||||||||||||||||||||||||||||||||||||||||||||
2451 CAAAGGCGGCTTCCAGTGTCTCTGCGCGGACCCCTACGAGTTAGGAGACG 2500

 834 spGlyArgThrCysValAspSerGlyArgLeuProArgValThr 848
     |||||||||||||||||||||||||||||||||||||:::|||
2501 ATGGGAGAACCTGCGTAGACTCCGGGAGGCTCCCTCGGGCGACT 2544
```

seq_name: gb_pr:HSASTPO2

```
seq_documentation_block:
LOCUS       HSASTPO2      2845 bp    mRNA            PRI       12-SEP-1993
DEFINITION  Human mRNA from alternatively spliced TPO-2 gene for thyroid
            peroxidase (EC.1.11.1.7).
ACCESSION   X17358
VERSION     X17358.1  GI:28897
KEYWORDS    autoantigen; peroxidase; thyroid peroxidase; TPO-2 gene;
            transmembrane protein.
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 2845)
  AUTHORS   Barnett,P.S.
  TITLE     Direct Submission
  JOURNAL   Submitted (20-DEC-1989) Barnett P.S., King's College School of
            Medicine & Dentistry, Denmark Hill, London SE5 8RX, United Kingdom
REFERENCE   2  (bases 1 to 2845)
  AUTHORS   Barnett,P.S., Banga,J.P., Watkins,J., Huang,G.C., Gluckman,D.R.,
            Page,M.J. and McGregor,A.M.
  TITLE     Nucleotide sequence of the alternatively spliced human thyroid
            peroxidase cDNA, TPO-2
  JOURNAL   Nucleic Acids Res. 18 (3), 670 (1990)
  MEDLINE   90175008
FEATURES             Location/Qualifiers
     source          1..2845
                     /organism="Homo sapiens"
                     /strain="Graves' patient"
                     /db_xref="taxon:9606"
                     /chromosome="2"
                     /tissue_type="thyroid"
                     /clone_lib="lambda gt10"
                     /clone="TPO-21"
     CDS             66..2696
                     /note="thyroid peroxidase (AA 1-876)"
                     /codon_start=1
                     /protein_id="CAA35235.1"
                     /db_xref="GI:28898"
                     /db_xref="SWISS-PROT:P07202"
                     /translation="MRALAVLSVTLVMACTEAFFPFISRGKELLWGKPEESRVSSVLE
                    ESKRLVDTAMYATMQRNLKKRGILSPAQLLSFSKLPEPTSGVIARAAEIMETSIQAMK
                    RKVNLKTQQSQHPTDALSEDLLSIIANMSGCLPYMLPPKCPNTCLANKYRPITGACNN
                    RDHPRWGASNTALARWLPPVYEDGFSQPRGWNPGFLYNGFPLPPVREVTRHVIQVSNE
                    VVTDDDRYSDLLMAWGQYIDHDIAFTPQSTSKAAFGGGADCQMTCENQNPCFPIQLPE
                    EARPAAGTACLPFYRSSAACGTGDQGALFGNLSTANPRQQMNGLTSFLDASTVYGSSP
                    ALERQLRNWTSAEGLLRVHARLRDSGRAYLPFVPPRAPSACAPEPGIPGETRGPCFLA
                    GDGRATEVPSLTALHTLWLREHNRLAAALKALNAHWSADAVYQEARKVVGALHQIITL
                    RDYIPRILGPEAFQQYVGPYEGYDSTANPTVSNVFSTAAFRFGHATIHPLVRRLDASF
                    QEHPDLPGLWLHQAFFSPWTLLRGGYNEWREFCGLPRLETPADLSTAIASRSVADKIL
                    DLYKHPDNIDVWLGGLAENFLPRARTGPLFACLIGKQMKALRDGDWFWWENSHVFTDA
                    QRRELEKHSLSRVICDNTGLTRVPMDAFQVGKFPEDFESCDSITGMNLEAWRETFPQD
                    DKCGFPESVENGDFVHCEESGRRVLVYSCRHGYELQGREQLTCTQEGWDFQPPLCKDV
                    NECADGAHPPCHASARCRNTKGGFQCLCADPYELGDDGRTCVDSGRLPRVTWISMSLA
                    ALLIGGFAGLTSTVICRWTRTGTKSTLPISETGGGTPELRCGKHQAVGTSPQRAAAQD
                    SEQESAGMEGRDTHRLPRAL"
BASE COUNT      622 a    880 c    819 g    524 t
ORIGIN
```

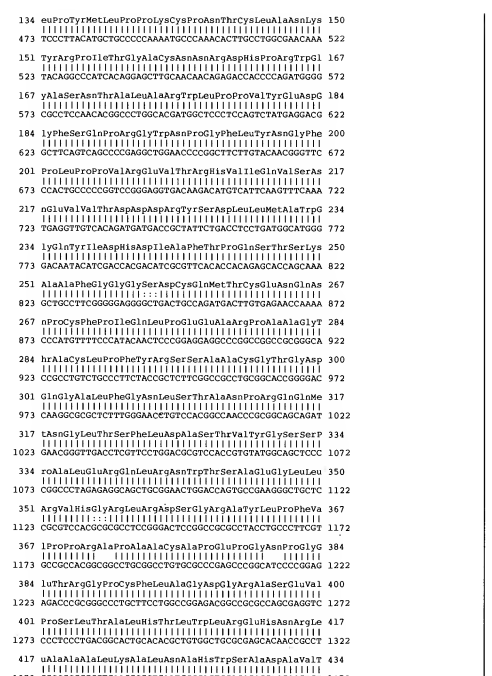
```
alignment_scores:
            Quality: 4215.50              Length:    848
              Ratio:  5.343                 Gaps:      1
 Percent Similarity:  93.042   Percent Identity: 92.571

alignment_block:
 US-08-482-402A-3_COPY_1_848 x HSASTPO2   ..

 Align seg 1/1  to: HSASTPO2  from: 1  to: 2845

   1 MetArgAlaLeuAlaValLeuSerValThrLeuValMetAlaCysThrGl 17
     ||||||||||||||||||||||||||||||||||||||||||||||||||
  66 ATGAGAGCGCTGGCTGTGCTGTCTGTCACGCTGGTTATGGCCTGCACAGA 115

  17 uAlaPhePheProPheIleSerArgGlyLysGluLeuLeuTrpGlyLysP 34
     ||||||||||||||||||||||||||||||||||||||||||||||||||
 116 AGCCTTCTTCCCCTTCATCTCGAGAGGGAAAGAACTCCTTTGGGGAAAGC 165

  34 roGluGluSerArgValSerSerValLeuGluGluSerLysArgLeuVal 50
     ||||||||||||||||||||||||||||||||||||||||||||||||||
 166 CTGAGGAGTCTCGTGTCTCTAGCGTCTTGGAGGAAAGCAAGCGCCTGGTG 215

  51 AspThrAlaMetTyrAlaThrMetGlnArgAsnLeuLysLysArgGlyIl 67
```



```
    667 yAspTrpPheTrpTrpGluAsnSerHisValPheThrAspAlaGlnArgA 684
        ||||||||||||||||||||||||||||||||||||||||||||||||||
   1895 TGACTGGTTTTGGTGGGAGAACAGCCACGTCTTCACGGATGCACAGAGGC 1944

    684 rgGluLeuGluLysHisSerLeuSerArgValIleCysAspAsnThrGly 700
        ||||||||||||||||||||||||||||||||||||||||||||||||||
   1945 GTGAGCTGGAGAAGCACTCCCTGTCTCGGGTCATCTGTGACAACACTGGC 1994

    701 LeuThrArgValProMetAspAlaPheGlnValGlyLysPheProGluAs 717
        ||||||||||||||||||||||||||||||||||||||||||||||||||
   1995 CTCACCAGGGTGCCCATGGATGCCTTCCAAGTCGGCAAATTCCCCGAAGA 2044

    717 pPheGluSerCysAspSerIleThrGlyMetAsnLeuGluAlaTrpArgG 734
        ||||||||||||||||||||||||||||||||||||||||||||||||||
   2045 CTTTGAGTCTTGTGACAGCATCACTGGCATGAACCTGGAGGCCTGGAGGG 2094

    734 luThrPheProGlnAspAspLysCysGlyPheProGluSerValGluAsn 750
        ||||||||||||||||||||||||||||||||||||||||||||||||||
   2095 AAACCTTTCCTCAAGACGACAAGTGTGGCTTCCCAGAGAGCGTGGAGAAT 2144

    751 GlyAspPheValHisCysGluGluSerGlyArgArgValLeuValTyrSe 767
        ||||||||||||||||||||||||||||||||||||||||||||||||||
   2145 GGGGACTTTGTGCACTGTGAGGAGTCTGGGAGGCGCGTGCTGGTGTATTC 2194

    767 rCysArgHisGlyTyrGluLeuGlnGlyArgGluGlnLeuThrCysThrG 784
        ||||||||||||||||||||||||||||||||||||||||||||||||||
   2195 CTGCCGGCACGGGTATGAGCTCCAAGGCCGGGAGCAGCTCACTTGCACCC 2244

    784 lnGluGlyTrpAspPheGlnProProLeuCysLysAspValAsnGluCys 800
        ||||||||||||||||||||||||||||||||||||||||||||||||||
   2245 AGGAAGGATGGGATTTCCAGCCTCCCCTCTGCAAAGATGTGAACGAGTGT 2294

    801 AlaAspGlyAlaHisProProCysHisAlaSerAlaArgCysArgAsnTh 817
        ||||||||||||||||||||||||||||||||||||||||||||||||||
   2295 GCAGACGGTGCCCACCCCCCCTGCCACGCCTCTGCGAGGTGCAGAAACAC 2344

    817 rLysGlyGlyPheGlnCysLeuCysAlaAspProTyrGluLeuGlyAspA 834
        ||||||||||||||||||||||||||||||||||||||||||||||||||
   2345 CAAAGGCGGCTTCCAGTGTCTCTGCGCGGACCCCTACGAGTTAGGAGACG 2394

    834 spGlyArgThrCysValAspSerGlyArgLeuProArgValThr 848
        ||||||||||||||||||||||||||||||||||||||||||||
   2395 ATGGGAGAACCTGCGTAGACTCCGGGAGGCTCCCTCGGGTGACT 2438
```

```
seq_name: gb_pr:HUMTPOB

seq_documentation_block:
LOCUS       HUMTPOB      2877 bp    mRNA            PRI       14-JAN-1995
DEFINITION  Human thyroid peroxidase mRNA, clone phTPO-2.4.
ACCESSION   J02970
VERSION     J02970.1  GI:339870
KEYWORDS    alternative splicing; thyroid peroxidase.
SOURCE      Human thyroid gland (from patients with Grave's disease), cDNA to
            mRNA, clone phTPO-2.4.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 2877)
  AUTHORS   Kimura,S., Kotani,T., McBride,O.W., Umeki,K., Hirai,K., Nakayama,T.
            and Ohtaki,S.
  TITLE     Human thyroid peroxidase: complete cDNA and protein sequence,
            chromosome mapping, and identification of two alternately spliced
            mRNAs
  JOURNAL   Proc. Natl. Acad. Sci. U.S.A. 84 (16), 5555-5559 (1987)
  MEDLINE   87289643
COMMENT     Computer-readable copy of sequence in [1] kindly provided by
            S.Kimura, 09-SEP-1987.
FEATURES             Location/Qualifiers
     source          1..2877
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /map="2pter-p24"
     mRNA            <1..2877
                     /note="TPO mRNA"
     gene            73..2703
                     /gene="TPO"
     CDS             73..2703
                     /gene="TPO"
                     /note="thyroid peroxidase"
                     /codon_start=1
                     /db_xref="GDB:G00-120-446"
                     /protein_id="AAA61216.1"
                     /db_xref="GI:339871"
                     /translation="MRALAVLSVTLVMACTEAFFPFISRGKELLWGKPEESRVSSVLE
                     ESKRLVDTAMYATMQRNLKKRGILSPAQLLSFSKLPEPTSGVIARAAEIMETSIQAMK
                     RKVNLKTQQSQHPTDALSEDLLSIIANMSGCLPYMLPPKCPNTCLANKYRPITGACNN
                     RDHPRWGASNTALARWLPPVYEDGFSQPRGWNPGFLYNGFPLPPVREVTRHVIQVSNE
                     VVTDDDRYSDLLMAWGQYIDHDIAFTPQSTSKAAFGGGADCQMTCENQNPCFPIQLPE
                     EARPAAGTACLPFYRSSAACGTGDQGALFGNLSTANPRQQMNGLTSFLDASTVYGSSP
                     ALERQLRNWTSAEGLLRVHARLRDSGRAYLPFVPPRRPAACAPEPGIPGETRGPCFLA
                     GDGRASEVPSLTALHTLWLREHNRLAAALKALNAHWSADAVYQEARKVVGALHQIITL
                     RDYIPRILGPEAFQQYVGPYEGYDSTANPTVSNVFSTAAFRFGHATIHPLVRRLDASF
                     QEHPDLPGLWLHQAFFSPWTLLRGGYNEWREFCGLPRLETPADLSTAIASRSVADKIL
                     DLYKHPDNIDVWLGGLAENFLPRARTGPLFACLIGKQMKALRDGDWFWWENSHVFTDA
                     QRRELEKHSLSRVICDNTGLTRVPMDAFQVGKFPEDFESCDSIPGMNLEAWRETFPQD
                     DKCGFPESVENGDFVHCEESGRRVLVYSCRHGYELQGREQLTCTQEGWDFQPPLCKDV
                     NECADGAHPPCHASARCRNTKGGFQCLCADPYELGDDGRTCVDSGRLPRATWISMSLA
                     ALLIGGFAGLTSTVICRWTRTGTKSTLPISETGGGTPELRCGKHQAVGTSPQRAAAQD
                     SEQESAGMEGRDTHRLPRAL"
BASE COUNT      632 a    883 c    829 g    533 t
ORIGIN      Chromosome 2pter-q11.
```

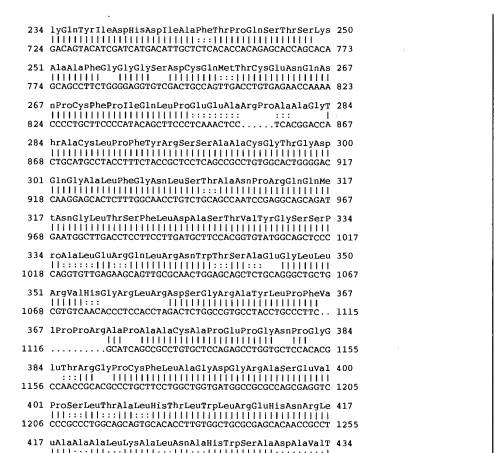
```
alignment_scores:
          Quality: 4206.50             Length:     848
            Ratio:  5.345                Gaps:       1
 Percent Similarity:  92.807   Percent Identity: 92.453

alignment_block:
 US-08-482-402A-3_COPY_1_848 x HUMTPOB   ..

 Align seg 1/1  to: HUMTPOB  from: 1  to: 2877

      1 MetArgAlaLeuAlaValLeuSerValThrLeuValMetAlaCysThrGl 17
        ||||||||||||||||||||||||||||||||||||||||||||||||||
     73 ATGAGAGCGCTCGCTGTGCTGTCTGTCACGCTGGTTATGGCCTGCACAGA 122

     17 uAlaPhePheProPheIleSerArgGlyLysGluLeuLeuTrpGlyLysP 34
        ||||||||||||||||||||||||||||||||||||||||||||||||||
    123 AGCCTTCTTCCCCTTCATCTCGAGAGGGAAAGAACTCCTTTGGGGAAAGC 172

     34 roGluGluSerArgValSerSerValLeuGluGluSerLysArgLeuVal 50
        ||||||||||||||||||||||||||||||||||||||||||||||||||
    173 CTGAGGAGTCTCGTGTCTCTAGCGTCTTGGAGGAAAGCAAGCGCCTGGTG 222

     51 AspThrAlaMetTyrAlaThrMetGlnArgAsnLeuLysLysArgGlyIl 67
        ||||||||||||||||||||||||||||||||||||||||||||||||||
    223 GACACCGCCATGTACGCCACGATGCAGAGAAACCTCAAGAAAAGAGGAAT 272

     67 eLeuSerGlyAlaGlnLeuLeuSerPheSerLysLeuProGluProThrS 84
        |||||||    |||||||||||||||||||||||||||||||||||||||
    273 CCTTTCTCCAGCTCAGCTTCTGTCTTTTTCCAAACTTCCTGAGCCAACAA 322

     84 erGlyValIleAlaArgAlaAlaGluIleMetGluThrSerIleGlnAla 100
        ||||||||||||||||||||||||||||||||||||||||||||||||||
    323 GCGGAGTGATTGCCCGAGCAGCAGAGATAATGGAAACATCAATACAAGCG 372

    101 MetLysArgLysValAsnLeuLysThrGlnGlnSerGlnHisProThrAs 117
        ||||||||||||||||||||||||||||||||||||||||||||||||||
    373 ATGAAAAGAAAAGTCAACCTGAAAACTCAACAATCACAGCATCCAACGGA 422

    117 pAlaLeuSerGluAspLeuLeuSerIleIleAlaAsnMetSerGlyCysL 134
        ||||||||||||||||||||||||||||||||||||||||||||||||||
    423 TGCTTTATCAGAAGATCTGCTGAGCATCATTGCAAACATGTCTGGATGTC 472
```

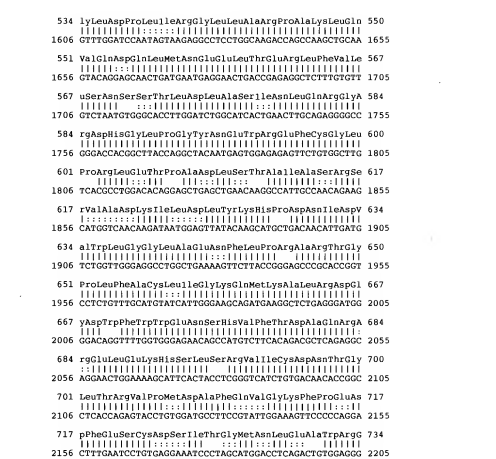

134 euProTyrMetLeuProProLysCysProAsnThrCysLeuAlaAsnLys 150
473 TCCCTTACATGCTGCCCCAAAATGCCCAAACACTTGCCTGGCGAACAAA 522
151 TyrArgProIleThrGlyAlaCysAsnAsnArgAspHisProArgTrpGl 167
523 TACAGGCCCATCACAGGAGCTTGCAACAACAGAGACCACCCCAGATGGGG 572
167 yAlaSerAsnThrAlaLeuAlaArgTrpLeuProProValTyrGluAspG 184
573 CGCCTCCAACACGGCCCTGGCACGATGGCTCCCTCCAGTCTATGAGGACG 622
184 lyPheSerGlnProArgGlyTrpAsnProGlyPheLeuTyrAsnGlyPhe 200
623 GCTTCAGTCAGCCCGGAGGCTGGAACCCCGGCTTCTTGTACAACGGGTTC 672
201 ProLeuProProValArgGluValThrArgHisValIleGlnValSerAs 217
673 CCACTGCCCCGGTCCGGGAGGTGACAAGACATGTCATTCAAGTTTCAAA 722
217 nGluValValThrAspAspAspArgTyrSerAspLeuLeuMetAlaTrpG 234
723 TGAGGTTGTCACAGATGATGACCGCTATTCTGACCTCCTGATGGCATGGG 772
234 lyGlnTyrIleAspHisAspIleAlaPheThrProGlnSerThrSerLys 250
773 GACAATACATCGACCACGACATCGCGTTCACACCACAGAGCACCAGCAAA 822
251 AlaAlaPheGlyGlyGlySerAspCysGlnMetThrCysGluAsnGlnAs 267
823 GCTGCCTTCGGGGGAGGGCTGACTGCCAGATGACTTGTGAGAACCAAAA 872
267 nProCysPheProIleGlnLeuProGluGluAlaArgProAlaAlaGlyT 284
873 CCCATGTTTTCCCATACAACTCCCGGAGGAGGCCCGGCCGGCCGCGGGCA 922
284 hrAlaCysLeuProPheTyrArgSerSerAlaAlaCysGlyThrGlyAsp 300
923 CCGCCTGTCTGCCCTTCTACCGCTCTTCGGCCGCCTGCGGCACCGGGGAC 972
301 GlnGlyAlaLeuPheGlyAsnLeuSerThrAlaAsnProArgGlnGlnMe 317
973 CAAGGCGCGCTTCTTGGGAACTGTCCACGGCCAACCCGCGGCAGCAGAT 1022
317 tAsnGlyLeuThrSerPheLeuAspAlaSerThrValTyrGlySerSerP 334
1023 GAACGGGTTGACCTCGTTCCTGGACGCGTCCACCGTGTATGGCAGCTCCC 1072
334 roAlaLeuGluArgGlnLeuArgAsnTrpThrSerAlaGluGlyLeuLeu 350
1073 CGGCCCTAGAGAGGCAGCTGCGGAACTGGACCAGTGCCGAAGGGCTGCTC 1122
351 ArgValHisGlyArgLeuArgAspSerGlyArgAlaTyrLeuProPheVa 367
1123 CGCGTCCACGCGCGCCTCCGGGACTCCGGCCGCCGCCTACCTGCCCTTCGT 1172
367 lProProArgAlaProAlaAlaCysAlaProGluProGlyAsnProGlyG 384
1173 GCCGCCACGGCGGCCTGCGGCCTGTGCGCCCGAGCCCGGCATCCCCGGAG 1222
384 luThrArgGlyProCysPheLeuAlaGlyAspGlyArgAlaSerGluVal 400
1223 AGACCCGCGGGCCCTGCTTCCTGGCCGGAGACGGCCGCGCCAGCGAGGTC 1272
401 ProSerLeuThrAlaLeuHisThrLeuTrpLeuArgGluHisAsnArgLe 417
1273 CCCTCCCTGACGGCACTGCACACGCTGTGGCTGCGCGAGCACAACCGCCT 1322
417 uAlaAlaAlaLeuLysAlaLeuAsnAlaHisTrpSerAlaAspAlaValT 434
1323 GGCCGCGGCGCTCAAGGCCCTCAATGCGCACTGGAGCGCGGACGCCGTGT 1372
434 yrGlnGluAlaArgLysValValGlyAlaLeuHisGlnIleIleThrLeu 450
1373 ACCAGGAGGCGCGCAAGGTCGTGGGCGCTCTGCACCAGATCATCACCCTG 1422
451 ArgAspTyrIleProArgIleLeuGlyProGluAlaPheGlnGlnTyrVa 467
1423 AGGGATTACATCCCCAGGATCCTGGGACCCGAGCCCTTCCAGCAGTACGT 1472
467 lGlyProTyrGluGlyTyrAspSerThrAlaAsnProThrValSerAsnV 484
1473 GGGTCCCTATGAAGGCTATGACTCCACCGCCAACCCCACTGTGTCCAACG 1522
484 alPheSerThrAlaAlaPheArgPheGlyHisAlaThrIleHisProLeu 500
1523 TGTTCTCCACAGCCGCCTTCCGCTTCGGCCATGCCACGATCCACCCGCTG 1572
501 ValArgArgLeuAspAlaSerPheGlnGluHisProAspLeuProGlyLe 517
1573 GTGAGGAGGCTGGACGCCAGTTCCAGGAGCACCCCGACCTGCCCGGGCT 1622
517 uTrpLeuHisGlnAlaPhePheSerProTrpThrLeuLeuArgGlyGlyG 534
1623 GTGGCTGCACCAGGCTTTCTTCAGCCCATGGACATTACTCCGTGGA.... 1668
534 lyLeuAspProLeuIleArgGlyLeuLeuAlaArgProAlaLysLeuGln 550
1668 1668
551 ValGlnAspGlnLeuMetAsnGluGluLeuThrGluArgLeuPheValLe 567
1668 1668
567 uSerAsnSerSerThrLeuAspLeuAlaSerIleAsnLeuGlnArgGlyA 584
1668 1668
584 rgAspHisGlyLeuProGlyTyrAsnGluTrpArgGluPheCysGlyLeu 600
1669 .....GGTTACAATGAGTGGAGGGAGTTCTGCGGCCTG 1701
601 ProArgLeuGluThrProAlaAspLeuSerThrAlaIleAlaSerArgSe 617
1702 CCTCGCCTGGAGACCCCGCTGACCTGAGCACAGCCATCGCCAGCAGGAG 1751
617 rValAlaAspLysIleLeuAspLeuTyrLysHisProAspAsnIleAspV 634
1752 CGTGGCCGACAAGATCCTGGACTTGTACAAGCATCCTGACAACATCGATG 1801
634 alTrpLeuGlyGlyLeuAlaGlyAsnPheLeuProArgAlaArgThrGly 650
1802 TCTGGCTGGGAGGCTTAGCTGAAAACTTCCTCCCCAGGGCTCGGACAGGG 1851
651 ProLeuPheAlaCysLeuIleGlyLysGlnMetLysAlaLeuArgAspGl 667
1852 CCCCTGTTTGCCTGTCTCATTGGGAAGCAGATGAAGGCTCTGCGGGATGG 1901
667 yAspTrpPheTrpTrpGluAsnSerHisValPheThrAspAlaGlnArgA 684
1902 TGACTGGTTTTGGTGGGAGAACAGCCACGTCTTCACCGATGCACAGAGGC 1951
684 rgGluLeuGluLysHisSerLeuSerArgValIleCysAspAsnThrGly 700
1952 GTGAGCTGGAGAAGCACTCCCTGTCTCGGGTCATCTGTGACAACACTGGC 2001
701 LeuThrArgValProMetAspAlaPheGlnValGlyLysPheProGluAs 717
2002 CTCACCAGGGTGCCCATGGATGCCTTCCAAGTCGGCAAATTCCCTGAAGA 2051
717 pPheGluSerCysAspSerIleThrGlyMetAsnLeuGluAlaTrpArgG 734
2052 CTTTGAGTCTTGTGACAGCATCCCTGCATGAACCTGGAGGCCTGGAGGG 2101
734 luThrPheProGlnAspAspLysCysGlyPhePheProGluSerValGluAsn 750

```
2102 AAACCTTTCCTCAAGACGACAAGTGTGGCTTCCCAGAGAGCGTGGAGAAT 2151

 751 GlyAspPheValHisCysGluGluSerGlyArgArgValLeuValTyrSe 767
     ||||||||||||||||||||||||||||||||||||||||||||||||||
2152 GGGGACTTTGTGCACTGTGAGGAGTCTGGGAGGCGCGTGCTGGTGTATTC 2201

 767 rCysArgHisGlyTyrGluLeuGlnGlyArgGluGlnLeuThrCysThrG 784
     ||||||||||||||||||||||||||||||||||||||||||||||||||
2202 CTGCCGGCACGGGTATGAGCTCCAAGGCCGGGAGCAGCTCACTTGCACCC 2251

 784 lnGluGlyTrpAspPheGlnProProLeuCysLysAspValAsnGluCys 800
     ||||||||||||||||||||||||||||||||||||||||||||||||||
2252 AGGAAGGATGGGATTTCCAGCCTCCCCTCTGCAAAGATGTGAACGAGTGT 2301

 801 AlaAspGlyAlaHisProProCysHisAlaSerAlaArgCysArgAsnTh 817
     ||||||||||||||||||||||||||||||||||||||||||||||||||
2302 GCAGACGGTGCCCACCCCCCCTGCCACGCCTCTGCGAGGTGCAGAAACAC 2351

 817 rLysGlyGlyPheGlnCysLeuCysAlaAspProTyrGluLeuGlyAspA 834
     ||||||||||||||||||||||||||||||||||||||||||||||||||
2352 CAAAGGCGGCTTCCAGTGTCTCTGCGCGGACCCCTACGAGTTAGGAGACG 2401

 834 spGlyArgThrCysValAspSerGlyArgLeuProArgValThr 848
     |||||||||||||||||||||||||||||||||||||||:::|||
2402 ATGGGAGAACCTGCGTAGACTCCGGGAGGCTCCCTCGGGCGACT 2445
```

seq_name: gb_ro:RNTPO

```
seq_documentation_block:
LOCUS       RNTPO       3237 bp    mRNA            ROD       12-SEP-1993
DEFINITION  Rat mRNA for thyroid peroxidase.
ACCESSION   X17396 M27275
VERSION     X17396.1  GI:57382
KEYWORDS    thyroid peroxidase.
SOURCE      Norway rat.
  ORGANISM  Rattus norvegicus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae;
            Rattus.
REFERENCE   1  (bases 1 to 3237)
  AUTHORS   Rapoport,B.
  TITLE     Direct Submission
  JOURNAL   Submitted (22-AUG-1989) to the EMBL/GenBank/DDBJ databases
REFERENCE   2  (bases 1 to 3237)
  AUTHORS   Derwahl,M., Seto,P. and Rapoport,B.
  TITLE     Complete nucleotide sequence of the cDNA for thyroid peroxidase in
            FRTL5 rat thyroid cells
  JOURNAL   Nucleic Acids Res. 17 (20), 8380 (1989)
  MEDLINE   90045972
FEATURES             Location/Qualifiers
     source          1..3237
                     /organism="Rattus norvegicus"
                     /db_xref="taxon:10116"
                     /cell_line="FRTL5."
     CDS             42..2786
                     /note="thyroid peroxidase (AA 1-914)"
                     /codon_start=1
                     /protein_id="CAA35257.1"
                     /db_xref="GI:57383"
                     /db_xref="SWISS-PROT:P14650"
                     /translation="MRTLGAMAVMLVVMGTAIFLPFLLRSRDILGGKTMTSHVISVVE
                     TSQLLVDNAVYNTMKRNLKKRGVLSPAQLLSFSKLPESTSGAISRAAEIMETSIQVMK
                     REQSQFSTDALSADILATIANLSGCLPFMLPPRCPDTCLANKYRPITGVCNNRDHPRW
                     GASNTALARWLPPVYEDGFSQPRGWNPNFLYHGFPLPPVREVTRHLIQVSNEAVTEDD
                     QYSDFLPVWGQYIDHDIALTPQSTSTAAFWGGVDCQLTCENQNPCFPIQLPSNSSRTT
                     ACLPFYRSSAACGTGDQGALFGNLSAANPRQQMNGLTSFLDASTVYGSSPGVEKQLRN
                     WSSSAGLLRVNTLHLDSGRAYLPFASAACAPEPGAPHANRTPCFLAGDGRASEVPALA
                     AVHTLWLREHNRLATAFKAINTHWSANTAYQEARKVVGALHQIITMRDYIPKILGPDA
                     FRQYVGPYEGYNPTVNPTVSNVFSTAAFRFGHATVHPLVRRLNTDFQDHTELPRLQLH
                     DVFFRPWRLIQEGGLDPIVRGLLARPAKLQVQEQLMNEELTERLFVLSNVGTLDLASL
                     NLQRGRDHGLPGYNEWREFCGLSRLDTGAELNKAIANRSMVNKIMELYKHADNIDVWL
                     GGLAEKFLPGARTGPLFACIIGKQMKALRDGDRFWWENSHVFTDAQRQELEKHSLPRV
                     ICDNTGLTRVPVDAFRIGKFPQDFESCEEIPSMDLRLWRETFPQDDKCVFPEKVDNGN
                     FVHCEESGKLVLVYSCFHGYKLQGQEQVTCTQNGWDSEPPVCKDVNECADLTHPPCHS
                     SAKCKNTKGSFQCVCTDPYMLGEDEKTCIDSGRLPRASWVSIALGALLIGGLASLSWT
                     VICRWTHADKKSTLLITERVTMESGFRKSQESGISPQKAEVQDAEQEPAYGSRVLLCE
                     "
BASE COUNT      816 a    874 c    831 g    716 t
ORIGIN

alignment_scores:
            Quality: 3563.00            Length:     848
              Ratio:   4.591              Gaps:       3
 Percent Similarity:  91.509   Percent Identity:  76.651

alignment_block:
 US-08-482-402A-3_COPY_1_848 x RNTPO  ..

 Align seg 1/1  to: RNTPO  from: 1  to: 3237

   1 MetArgAlaLeuAlaValLeuSerValThrLeuValMetAlaCysThrGl 17
     ||||||:::|||:::::::::::|||    |||||:::       |||
  42 ATGAGAACACTTGGAGCTATGGCAGTAATGCTGGTTGTGATGGGAACTGC 91

  17 uAlaPhePheProPheIleSerArgGlyLysGluLeuLeuTrpGlyLysP 34
         |||:::||||||:::    |||:::::::::::|||    ||||||
  92 AATTTTCCTCCCTTTCCTCCTGAGAAGCAGAGACATCTTGGGGGGGAAGA 141

  34 roGluGluSerArgValSerSerValLeuGluGluSerLysArgLeuVal 50
             |||:::|||    ||||||:::|||    |||:::   ||||||
 142 CTATGACATCCCATGTTATCAGTGTTGTAGAAACGAGCCAGCTTCTGGTG 191

  51 AspThrAlaMetTyrAlaThrMetGlnArgAsnLeuLysLysArgGlyIl 67
     |||:::|||:::|||    ||||||:::|||||||||||||||||||||::
 192 GACAATGCAGTCTACAACACCATGAAAAGAAACCTCAAGAAAAGGGGAGT 241

  67 eLeuSerGlyAlaGlnLeuLeuSerPheSerLysLeuProGluProThrS 84
     :||||||   |||||||||||||||||||||||||||||||    ||||
 242 CCTTTCTCCAGCCCAGCTTCTCTCTTTCTCCAAGCTGCCAGAGTCCACCA 291

  84 erGlyValIleAlaArgAlaAlaGluIleMetGluThrSerIleGlnAla 100
     ||||||:::|||:::||||||||||||||||||||||||||||||||:::
 292 GTGGAGCTATTTCCCGAGCAGCAGAGATCATGGAAACATCAATACAAGTC 341

 101 MetLysArgLysValAsnLeuLysThrGlnGlnSerGlnHisProThrAs 117
     |||||||||                 :::||||||||||      |||||
 342 ATGAAACGT.................GAACAATCACAGTTCTCCACGGA 373

 117 pAlaLeuSerGluAspLeuLeuSerIleIleAlaAsnMetSerGlyCysL 134
     ||||||||||    |||:::|||:::    |||||||||:::||||||||||
 374 TGCACTATCAGCAGACATCCTGGCCACAATTGCCAACCTGTCAGGATGCT 423

 134 euProTyrMetLeuProProLysCysProAsnThrCysLeuAlaAsnLys 150
     ||||||:::|||||||||||||:::||||||:::||||||||||||||||
 424 TGCCTTTCATGCTGCCACCGAGATGTCCTGACACCTGCCTGGCAAACAAG 473

 151 TyrArgProIleThrGlyAlaCysAsnAsnArgAspHisProArgTrpGl 167
     ||||||||||||||||||:::|||||||||||||||||||||||||||||
 474 TACCGGCCCATCACAGGGGTGTGCAACAACAGAGATCACCCCAGATGGGG 523

 167 yAlaSerAsnThrAlaLeuAlaArgTrpLeuProProValTyrGluAspG 184
     ||||||||||||||||||||||||||||||||||||||||||||||||||
 524 AGCCTCCAACACAGCCCTAGCAAGATGGCTGCCTCCTGTCTACGAAGATG 573

 184 lyPheSerGlnProArgGlyTrpAsnProGlyPheLeuTyrAsnGlyPhe 200
     ||||||||||||||||||||||||||||:::|||||||||:::||||||
 574 GTTTCAGTCAGCCCAGAGGCTGGAACCCTAATTTCTTATACCATGGCTTC 623

 201 ProLeuProProValArgGluValThrArgHisValIleGlnValSerAs 217
     ||||||||||||||||||||||||||||||||||||:::||||||||||||
 624 CCACTGCCCCCGGTACGGGAGGTGACAAGACACCTCATTCAAGTTTCAAA 673

 217 nGluValValThrAspAspAspArgTyrSerAspLeuLeuMetAlaTrpG 234
     ||||:::||||||:::||||||:::|||||||||:::|||    :::||||
 674 TGAGGCTGTGACTGAAGATGACCAGTACTCTGATTTTCTGCCGGTGTGGG 723
```


234 lyGlnTyrIleAspHisAspIleAlaPheThrProGlnSerThrSerLys 250
724 GACAGTACATCGATCATGACATTGCTCTCACACCACAGAGCACCAGCACA 773
251 AlaAlaPheGlyGlyGlySerAspCysGlnMetThrCysGluAsnGlnAs 267
774 GCAGCCTTCTGGGGAGGTGTCGACTGCCAGTTGACCTGTGAGAACCAAAA 823
267 nProCysPheProIleGlnLeuProGluGluAlaArgProAlaAlaGlyT 284
824 CCCCTGCTTCCCCATACAGCTTCCCTCAAACTCC.....TCACGGACCA 867
284 hrAlaCysLeuProPheTyrArgSerSerAlaAlaCysGlyThrGlyAsp 300
868 CTGCATGCCTACCTTTCTACCGCTCCTCAGCCGCCTGTGGCACTGGGGAC 917
301 GlnGlyAlaLeuPheGlyAsnLeuSerThrAlaAsnProArgGlnGlnMe 317
918 CAAGGAGCACTCTTTGGCAACCTGTCTGCAGCCAATCCGAGGCAGCAGAT 967
317 tAsnGlyLeuThrSerPheLeuAspAlaSerThrValTyrGlySerSerP 334
968 GAATGGCTTGACCTCCTTCCTTGATGCTTCCACGGTGTATGGCAGCTCCC 1017
334 roAlaLeuGluArgGlnLeuArgAsnTrpThrSerAlaGluGlyLeuLeu 350
1018 CAGGTGTTGAGAAGCAGTTGCGCAACTGGAGCAGCTCTGCAGGGCTGCTG 1067
351 ArgValHisGlyArgLeuArgAspSerGlyArgAlaTyrLeuProPheVa 367
1068 CGTGTCAACACCCTCCACCTAGACTCTGGCCGTGCCTACCTGCCCTTC.. 1115
367 lProProArgAlaProAlaAlaCysAlaProGluProGlyAsnProGlyG 384
1116 .....GCATCAGCGCGCCTGTGCTCCAGAGCCTGGTGCTCCACACG 1155
384 luThrArgGlyProCysPheLeuAlaGlyAspGlyArgAlaSerGluVal 400
1156 CCAACCGCACGCCTGCTTCCTGGCTGGTGATGGCCGCCCAGCGAGGTC 1205
401 ProSerLeuThrAlaLeuHisThrLeuTrpLeuArgGluHisAsnArgLe 417
1206 CCCGCCCTGGCAGCAGTGCACACCTTGTGGCTGCGCGAGCACAACCGCCT 1255
417 uAlaAlaAlaLeuLysAlaLeuAsnAlaHisTrpSerAlaAspAlaVal 434
1256 GGCTACAGCTTTCAAAGCATCAACACACATTGGAGCGCCAACACTGCCT 1305
434 yrGlnGluAlaArgLysValValGlyAlaLeuHisGlnIleTheThrLeu 450
1306 ACCAGGAGCACGCAAGGTGGTAGGGCCATTGCACCAGATCATCACCATG 1355
451 ArgAspTyrIleProArgIleLeuGlyProGluAlaPheGlnGlnTyrVa 467
1356 AGGGATTATATCCCCAAGATCCTGGCTCCCGATGCCTTCAGGCAGTATGT 1405
467 lGlyProTyrGluGlyTyrAspSerThrAlaAsnProThrValSerAsnV 484
1406 GGGCCCCTATGAAGCTATAACCCCACGGTGAACCCTACTGTGTCCAACG 1455
484 alPheSerThrAlaAlaPheArgPheGlyHisAlaThrIleHisProLeu 500
1456 TCTTCTCTACTGCTGCCTTCCGCTTTGGCCATGCCACAGTCCATCCACTG 1505
501 ValArgArgLeuAspAlaSerPheGlnGluHisProAspLeuProGlyLe 517
1506 GTGAGACGACTAAACACGGACTTCCAGGACCACACAGAGCTCCCCAGGCT 1555
517 uTrpLeuHisGlnAlaPhePheSerProTrpThrLeuLeuArgGlyGlyG 534
1556 GCAGCTGCACGACGTCTTCTTCAGGCCCTGGAGGCTTATCCAGGAAGGTG 1605


534 lyLeuAspProLeuIleArgGlyLeuLeuAlaArgProAlaLysLeuGln 550
1606 GTTTGGATCCAATAGTAAGAGGCCTCCTGGCAAGACCAGCCAAGCTGCAA 1655
551 ValGlnAspGlnLeuMetAsnGluGluLeuThrGluArgLeuPheValLe 567
1656 GTACAGGAGCAACTGATGAATGAGGAACTGACCGAGAGGCTCTTTGTGTT 1705
567 uSerAsnSerSerThrLeuAspLeuAlaSerIleAsnLeuGlnArgGlyA 584
1706 GTCTAATGTGGGCACCTTGGATCTGGCATCACTGAACTTGCAGAGGGGCC 1755
584 rgAspHisGlyLeuProGlyTyrAsnGluTrpArgGluPheCysGlyLeu 600
1756 GGGACCAGGCTTACCAGGCTACAATGAGTGGAGAGAGTTCTGTGGCTTG 1805
601 ProArgLeuGluThrProAlaAspLeuSerThrAlaIleAlaSerArgSe 617
1806 TCAGCCTGGACACAGGAGCTGAGCTGAACAAGGCCATTGCCAACAGAAG 1855
617 rValAlaAspLysIleLeuAspLeuTyrLysHisProAspAsnIleAspV 634
1856 CATGGTCAACAAGATAATGGAGTTATACAAGCATGCTGACAACATTGATG 1905
634 alTrpLeuGlyGlyLeuAlaGluAsnPheLeuProArgAlaArgThrGly 650
1906 TCTGGTTGGGAGGCCTGGCTGAAAAGTTCTTACCGGGAGCCCGCACCGGT 1955
651 ProLeuPheAlaCysLeuIleGlyLysGlnMetLysAlaLeuArgAspGl 667
1956 CCTCTGTTTGCATGTATCATTGGGAAGCAGATGAAGGCTCTGAGGGATGG 2005
667 yAspTrpPheTrpTrpGlnAsnSerHisValPheThrAspAlaGlnArgA 684
2006 GGACAGGTTTGGTGGGAGAACAGCCATGTCTTCACAGACGCTCAGAGGC 2055
684 rgGluLeuGluLysHisSerLeuSerArgValIleCysAspAsnThrGly 700
2056 AGGAACTGGAAAAGCATTCACTACCTCGGGTCATCTGTGACAACACCGGC 2105
701 LeuThrArgValProMetAspAlaPheGlnValGlyLysPheProGluAs 717
2106 CTCACCAGAGTACCTGTGGATGCCTTCCGTATTGGAAAGTTCCCCCAGGA 2155
717 pPheGluSerCysAspSerIleThrGlyMetAsnLeuGluAlaTrpArgG 734
2156 CTTTGAATCCTGTGAGGAAATCCCTAGCATGGACCTCAGACTGTGGAGGG 2205
734 luThrPheProGlnAspAspLysCysGlyPhePheGluSerValGluAsn 750
2206 AGACCTTCCCACAAGACGACAAGTGTGGTCTTCCAGAGAAGGTGGACAAT 2255
751 GlyAspPheValHisCysGluGluSerGlyArgArgValLeuvalTyrSe 767
2256 GGGAACTTTGTCACTGTGAAGAATCTGGGAAGCTGGTACTGGTGTATTC 2305
767 rCysArgHisGlyTyrGluLeuGlnGlyArgGluGlnLeuThrCysThrG 784
2306 CTGTTTCCATGGATACAAGCTGCAAGGCCAGGAGCAGGTCACATGTACCC 2355
784 lnGluGlyTrpAspPheGlnPheProLeuCysLysAspValAsnGluCys 800
2356 AGAATGGATGGCACTCAGAGCCTCCTGTCTGTAAAGATGTTAATGAGTGT 2405
801 AlaAspGlyAlaHisProProCysHisAlaSerAlaArgCysArgAsnTh 817
2406 GCAGATCTGACACACCCACCTGTCCACTCCTCCCGCAAAGTGCAAGAACAC 2455
817 rLysGlyGlyPheGlnCysLeuCysAlaAspProTyrGluLeuGlyAspA 834
2456 CAAGGGAAGCTTCCAGTGTGTGTGCACAGACCCCTACATGCTAGGTGAGG 2505
834 spGlyArgThrCysValAspSerGlyArgLeuProArgValThr 848

```
            ||    :::||||||:::||||||||||||||||||||||:::::::
      2506 ATGAGAAGACCTGCATAGATTCTGGCAGGCTACCTCGGGCATCC 2549

seq_name: gb_ro:MMTPER

seq_documentation_block:
LOCUS       MMTPER         3291 bp    mRNA            ROD       25-AUG-1999
DEFINITION  Mus musculus mRNA for thyroid peroxidase.
ACCESSION   X60703
VERSION     X60703.1  GI:297160
KEYWORDS    peroxidase.
SOURCE      house mouse.
  ORGANISM  Mus musculus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE   1  (bases 1 to 3291)
  AUTHORS   Kotani,T., Umeki,K., Yamamoto,I., Takeuchi,M., Takechi,S.,
            Nakayama,T. and Ohtaki,S.
  TITLE     Nucleotide sequence of the cDNA encoding mouse thyroid peroxidase
  JOURNAL   Gene 123 (2), 289-290 (1993)
  MEDLINE   93154601
REFERENCE   2  (bases 1 to 3291)
  AUTHORS   Sachiya,O.
  TITLE     Direct Submission
  JOURNAL   Submitted (26-JUN-1991) O. Sachiya, Central Laboratopy for Clinical
            Investigation, Medical College Hospital, Miyazaki Medical College,
            Kiyotake, Miyazaki 889-16, JAPAN
FEATURES             Location/Qualifiers
     source          1. .3291
                     /organism="Mus musculus"
                     /strain="C57BL/6"
                     /db_xref="taxon:10090"
                     /tissue_type="thyroid"
                     /cell_type="follicular"
                     /clone="pmTPO-75"
     mRNA            74. .3281
                     /partial
                     /note="peroxidase"
                     /evidence=experimental
     CDS             74. .2818
                     /EC_number="1.11.1.7"
                     /codon_start=1
                     /product="peroxidase"
                     /protein_id="CAA43114.1"
                     /db_xref="GI:4539541"
                     /db_xref="MGD:MGI:98813"
                     /db_xref="SWISS-PROT:P35419"
                     /translation="MRTLGAMAIMLVVMGTVIFLSFILRSRDILCGKTMKSHVISAVE
                     TSQLMVDHAVYNTMKRNLKKREVLSPAQLLSFFKLPESTSGAISRAAEIMETSIQVMK
                     REQSQFSTDALSADILGTIANLSGCLPFMLPPRCPDTCLANKYRPITGACNNRDHPRW
                     GASNTALARWLPPVYEDGFSQPKGWNPNFLYHGFPLPPVREVTRHLIQVSNEAVTEDD
                     QYSDFLPVWGQYIDHDIALTPQSTSTAAFWGGVDCQLTCENQNPCFPIQLPSNSSGTT
                     ACLPFYRSSAACGTGDQGALFGNLSAANPRQQMNGLTSFLDASTVYGSSPGVEKQLRN
                     WSSSAGLLRVNTLHLDAGRAYLPFATAACAPEPGTPRTNRTPCFLAGDGRASEVPALA
                     AVHTLWLREHNRLASAFKAINKHWSANTAYQEARKVVGALHQIITMRDYIPKILGPDA
                     FRQYVGPYEGYNPTVNPTVSNIFSTAAFRFGHATVHPLVRRLNTDFQEHTELPRLQLR
                     DVFFRPWRLIQEGGLDPIVRGLLARAAKLQVQGQLMNEELTERLFVLSNVGTLDLASL
                     NLQRGRDHGLPDYNEWREFCGLSRLETPAELNKAIANRSMVNKIMDLYKHADNIDVWL
                     GGLAEKFLPGARTGPLFACIIGKQMKALRDGDRFWWENTNVFTDAQRQELEKHSLPRV
                     ICDNTGLTRVPVDAFRIGKFPQDFESCEDIPSMDLELWRETFPQDDKCVFPEEVDNGN
                     FVHCEESGKLVLVYSCFHGYKLQGQEQVTCTQKGWDSEPPVCKDVNECADLTHPPCHP
                     SAQCKNTKGSFQCVCTDPYVLGEDEKTCIDSGRLPRASWVSIALGALLIGGLASLTWI
                     VICRWTHADKKATLPITERVTTQSGCRKSQGRGISPHKAAAQDTGQEPASGSRVLLCE
                     "
BASE COUNT      819 a    897 c    854 g    721 t
ORIGIN

alignment_scores:
          Quality: 3520.00             Length:     848
            Ratio:  4.565                Gaps:       3
 Percent Similarity:  90.920   Percent Identity: 75.825
```

```
alignment_block:
 US-08-482-402A-3_COPY_1_848 x MMTPER   ..

 Align seg 1/1   to: MMTPER   from: 1   to: 3291

      1 MetArgAlaLeuAlaValLeuSerValThrLeuValMetAlaCysThrGl 17
        ||||||:::|||::::::::::::::   |||||:::      |||
     74 ATGAGAACACTTGGAGCTATGGCAATAATGCTGGTGGTTATGGGAACTGT 123

     17 uAlaPhePheProPheIleSerArgGlyLysGluLeuLeuTrpGlyLysP 34
            |||:::   ||||||   |||::::::::::|||   ||||||
    124 AATTTTCCTCTCTTTTATCCTGAGAAGCAGAGACATCTTGTGTGGGAAGA 173

     34 roGluGluSerArgValSerSerValLeuGluGluSerLysArgLeuVal 50
          :::|||:::|||   |||::::::|||   |||:::    :::|||
    174 CCATGAAGTCCCATGTTATCAGTGCTGTGGAAACGAGCCAGCTCATGGTG 223

     51 AspThrAlaMetTyrAlaThrMetGlnArgAsnLeuLysLysArgGlyIl 67
        |||   |||:::|||   ||||||:::|||||||||||||||||   ::
    224 GACCATGCAGTCTACAACACCATGAAAAGAAACCTCAAGAAAAGGGAAGT 273

     67 eLeuSerGlyAlaGlnLeuLeuSerPheSerLysLeuProGluProThrS 84
        :||||||   |||||||||||||||||   ||||||||||||   ||||
    274 CCTTTCTCCAGCCCAGCTTCTCTCTTTCTTTAAGCTGCCCGAGTCCACCA 323

     84 erGlyValIleAlaArgAlaAlaGluIleMetGluThrSerIleGlnAla 100
        |||||:::|||:::|||||||||||||||||||||||||||||||||:::
    324 GTGGGGCTATTTCCCGAGCAGCAGAGATTATGGAAACATCAATACAAGTC 373

    101 MetLysArgLysValAsnLeuLysThrGlnGlnSerGlnHisProThrAs 117
        |||||||||                 :::|||||||||      |||||
    374 ATGAAACGT.................GAACAGTCACAGTTCTCCACGGA 405

    117 pAlaLeuSerGluAspLeuLeuSerIleIleAlaAsnMetSerGlyCysL 134
        ||||||||||   |||:::|||:::   |||||||||:::||||||||||
    406 TGCCTTATCAGCAGACATTCTGGGCACAATTGCCAACCTGTCAGGATGCT 455

    134 euProTyrMetLeuProProLysCysProAsnThrCysLeuAlaAsnLys 150
        |||||:::|||||||||||||:::||||||:::|||||||||||||||||
    456 TGCCTTTCATGCTGCCACCAAGATGTCCTGACACCTGCCTGGCAAATAAG 505

    151 TyrArgProIleThrGlyAlaCysAsnAsnArgAspHisProArgTrpGl 167
        ||||||||||||||||||||||||||||||||||||||||||||||||||
    506 TACCGGCCCATCACAGGGGCGTGCAACAATAGAGATCACCCCAGATGGGG 555

    167 yAlaSerAsnThrAlaLeuAlaArgTrpLeuProProValTyrGluAspG 184
        ||||||||||||||||||||||||||||||||||||||||||||||||||
    556 AGCCTCCAACACAGCCCTAGCAAGATGGCTGCCTCCTGTCTATGAAGATG 605

    184 lyPheSerGlnProArgGlyTrpAsnProGlyPheLeuTyrAsnGlyPhe 200
        ||||||||||||||:::|||||||||||||||||||||||:::||||||
    606 GCTTCAGTCAGCCCAAAGGCTGGAACCCTAATTTCTTATACCACGGCTTC 655

    201 ProLeuProProValArgGluValThrArgHisValIleGlnValSerAs 217
        ||||||||||||||||||||||||||||||:::|||||||||||||||||
    656 CCACTGCCCCCGGTACGGGAAGTGACAAGGCACCTCATTCAAGTTTCCAA 705

    217 nGluValValThrAspAspAspArgTyrSerAspLeuLeuMetAlaTrpG 234
        ||||:::||||||:::||||||:::||||||||||:::|||    :::||||
    706 TGAGGCTGTGACCGAAGATGACCAGTACTCTGATTTTCTGCCGGTGTGGG 755

    234 lyGlnTyrIleAspHisAspIleAlaPheThrProGlnSerThrSerLys 250
        |||||||||||||||||||||||||:::|||||||||||||||||
    756 GACAGTACATCGATCATGACATTGCTCTCACACCACAGAGCACTAGCACA 805

    251 AlaAlaPheGlyGlyGlySerAspCysGlnMetThrCysGluAsnGlnAs 267
        |||||||||   ||||||   |||||||||:::|||||||||||||||||
    806 GCAGCCTTCTGGGGAGGTGTCGACTGCCAGCTGACCTGTGAGAACCAAAA 855

    267 nProCysPheProIleGlnLeuProGluGluAlaArgProAlaAlaGlyT 284
        ||||||||||||||||||||||||||:::::::::      ::::::   |
    856 TCCTTGCTTCCCCATACAGCTTCCCTCAAACTCC......TCAGGGACCA 899
```


```
seq_name: gb_ro:RATTPOPR

seq_documentation_block:
LOCUS       RATTPOPR       2777 bp    mRNA            ROD       27-APR-1993
DEFINITION  Rat thyroid peroxidase (TPOP) mRNA, 3' end.
ACCESSION   M31655
VERSION     M31655.1  GI:207434
KEYWORDS    thyroid peroxidase.
SOURCE      Rat thyroid cell line FRTL-5, cDNA to mRNA.
  ORGANISM  Rattus norvegicus
```

```
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae;
            Rattus.
REFERENCE   1  (bases 1 to 2777)
  AUTHORS   Isozaki,O., Kohn,L.D., Kozak,C.A. and Kimura,S.
  TITLE     Thyroid peroxidase: Rat cDNA sequence, chromosomal localization in
            mouse, and regulation of gene expression by comparison to
            thyroglobulin in rat FRTL-5 cells
  JOURNAL   Mol. Endocrinol. 3, 1681-1692 (1989)
  MEDLINE   90114171
FEATURES             Location/Qualifiers
     source          1. .2777
                     /organism="Rattus norvegicus"
                     /db_xref="taxon:10116"
     CDS             <1. .2313
                     /note="thyroid peroxidase"
                     /codon_start=1
                     /protein_id="AAA42265.1"
                     /db_xref="GI:207435"
                     /translation="YRPITGVCNNRDHPRWGASNTALARWLPPVYEDGFSQPRGWNPN
                     FLYHGLGLPSVREVTRHLIQVSNEAVTEDDQYSDFLPVWAQYIDHDIALTPQSTSTAA
                     FWGGVDCQLTCENQNPCFPIQLPSNSSRTTACLPFYRSSAACGTGDQGALFGNLSAAN
                     PRQQMNGLTSFLDASTVYGSSPGVEKQLRNWSSSAGLLRVNTLHLDSGRAYLPFASAA
                     CAPEPGAPHANRTPCFLAGDGRASEVPALAAVHTLWLREHNRLATAFKAINTHWSANT
                     AYQEARKVVGALHQIITMRDYIPKILGPDAFRQYVGPYEGYNPTVNPTVSNVFSTAAF
                     RFGHATVHPLVRRLNTDFQDHTELPRLQLHDVFFRPWRLIQEGGLDPIVRGLLARPAK
                     LQVQEQLMNEELTERLFVLSNVGTLDLASLNLQRGRDHGLPGYNEWREFCGLSRLETP
                     AELNKAIANRSMVNKIMELYKHADNIDVWLGGLAEKFLPGARTGPLFACIIGKQMKAL
                     RDGDRFWWENSHVFTDAQRQELEKHSLPRVICDNTGLTRVPVDAFRIGKFPQDFESCE
                     EIPSMDLRLWRETFPQDDKCVFPEKVDNGNFVHCEESGKLVLVYSCFHGYKLQGQEQV
                     TCTQNGWDSEPPVCKDVNECADLTHPPCHSSAKCKNTKGSFQCVCTDPYMLGEDEKTC
                     IDSGRLPRASWVSIALGALLIGGLASLSWTVICRWTHADKKSTLLITERVTMESGFRK
                     SQESGISPQKAEVQDAEQEPAYGSRVLLCE"
BASE COUNT      690 a    752 c    722 g    613 t
ORIGIN


alignment_scores:
          Quality: 3082.00            Length:     698
            Ratio:   4.734              Gaps:       2
 Percent Similarity:  93.266   Percent Identity: 79.226

alignment_block:
 US-08-482-402A-3_COPY_1_848 x RATTPOFR   ..

 Align seg 1/1  to: RATTPOFR  from: 1  to: 2777

     151 TyrArgProIleThrGlyAlaCysAsnAsnArgAspHisProArgTrpGl 167
         ||||||||||||||||||:::||||||||||||||||||||||||||||||
       1 TACCGGCCCATCACAGGGGTGTGCAACAACAGAGATCACCCCAGATGGGG 50

     167 yAlaSerAsnThrAlaLeuAlaArgTrpLeuProProValTyrGluAspG 184
         ||||||||||||||||||||||||||||||||||||||||||||||||||
      51 AGCCTCCAACACAGCCCTAGCAAGATGGCTGCCTCCTGTCTACGAAGATG 100

     184 lyPheSerGlnProArgGlyTrpAsnProGlyPheLeuTyrAsnGlyPhe 200
         ||||||||||||||||||||||||||:::|||||||||:::|||:::
     101 GTTTCAGTCAGCCCAGAGGCTGGAACCCTAATTTCTTATACCATGGCTTG 150

     201 ProLeuProProValArgGluValThrArgHisValIleGlnValSerAs 217
           ||||||   ||||||||||||||||||||||:::||||||||||||||
     151 GGACTGCCCTCGGTACGGGAGGTGACAAGACACCTCATTCAAGTTTCAAA 200

     217 nGluValValThrAspAspAspArgTyrSerAspLeuLeuMetAlaTrpG 234
         ||||:::||||||:::||||||:::||||||||||:::|||    :::|||:
     201 TGAGGCTGTGACTGAAGATGACCAGTACTCTGATTTTCTGCCGGTGTGGG 250

     234 lyGlnTyrIleAspHisAspIleAlaPheThrProGlnSerThrSerLys 250
         :::||||||||||||||||||||||||:::|||||||||||||||||
     251 CACAGTACATCGATCATGACATTGCTCTCACACCACAGAGCACCAGCACA 300

     251 AlaAlaPheGlyGlyGlySerAspCysGlnMetThrCysGluAsnGlnAs 267
         |||||||||   ||||||   |||||||||:::|||||||||||||||

     301 GCAGCCTTCTGGGGAGGTGTCGACTGCCAGTTGACCTGTGAGAACCAAAA 350

     267 nProCysPheProIleGlnLeuProGluGluAlaArgProAlaAlaGlyT 284
         |||||||||||||||||||||||||:::::::::      :::      |
     351 CCCCTGCTTCCCCATACAGCTTCCCTCAAACTCC......TCACGGACCA 394

     284 hrAlaCysLeuProPheTyrArgSerSerAlaAlaCysGlyThrGlyAsp 300
         ||||||||||||||||||||||||||||||||||||||||||||||||||
     395 CTGCATGCCTACCTTTCTACCGCTCCTCAGCCGCCTGTGGCACTGGGGAC 444

     301 GlnGlyAlaLeuPheGlyAsnLeuSerThrAlaAsnProArgGlnGlnMe 317
         |||||||||||||||||||||||||||||:::||||||||||||||||||
     445 CAAGGAGCACTCTTTGGCAACCTGTCTGCAGCCAATCCGAGGCAGCAGAT 494

     317 tAsnGlyLeuThrSerPheLeuAspAlaSerThrValTyrGlySerSerP 334
         ||||||||||||||||||||||||||||||||||||||||||||||||||
     495 GAATGGCTTGACCTCCTTCCTTGATGCTTCCACGGTGTATGGCAGCTCCC 544

     334 roAlaLeuGluArgGlnLeuArgAsnTrpThrSerAlaGluGlyLeuLeu 350
         ||:::::|||:::||||||||||||||:::|||:::   |||||||||
     545 CAGGTGTTGAGAAGCAGTTGCGCAACTGGAGCAGCTCTGCAGGGCTGCTG 594

     351 ArgValHisGlyArgLeuArgAspSerGlyArgAlaTyrLeuProPheVa 367
         ||||||:::              ||||||||||||||||||||||||||
     595 CGTGTCAACACCCTCCACCTAGACTCTGGCCGTGCCTACCTGCCCTTC.. 642

     367 lProProArgAlaProAlaAlaCysAlaProGluProGlyAsnProGlyG 384
                   |||   ||||||||||||||||||||||||   |||
     643 .........GCATCAGCCGCCTGTGCTCCAGAGCCTGGTGCTCCACACG 682

     384 luThrArgGlyProCysPheLeuAlaGlyAspGlyArgAlaSerGluVal 400
           :::|||   |||||||||||||||||||||||||||||||||||||||
     683 CCAACCGCACGCCCTGCTTCCTGGCTGGTGATGGCCGCGCCAGCGAGGTC 732

     401 ProSerLeuThrAlaLeuHisThrLeuTrpLeuArgGluHisAsnArgLe 417
         |||:::|||:::|||:::||||||||||||||||||||||||||||||||
     733 CCCGCCCTGGCAGCAGTGCACACCTTGTGGCTGCGCGAGCACAACCGCCT 782

     417 uAlaAlaAlaLeuLysAlaLeuAsnAlaHisTrpSerAlaAspAlaValT 434
         ||||:::|||:::||||||:::|||:::||||||||||||:::::::::|
     783 GGCTACAGCTTTCAAAGCCATCAACACACATTGGAGCGCCAACACTGCCT 832

     434 yrGlnGluAlaArgLysValValGlyAlaLeuHisGlnIleIleThrLeu 450
         |||||||||||||||||||||||||||||||||||||||||||||||:::
     833 ACCAGGAGGCACGCAAGGTGGTAGGGGCATTGCACCAGATCATCACCATG 882

     451 ArgAspTyrIleProArgIleLeuGlyProGluAlaPheGlnGlnTyrVa 467
         ||||||||||||||||||:::||||||||||||||||||:::||||||||
     883 AGGGATTATATCCCCAAGATCCTGGGTCCCGATGCCTTCAGGCAGTATGT 932

     467 lGlyProTyrGluGlyTyrAspSerThrAlaAsnProThrValSerAsnV 484
         |||||||||||||||||||:::   |||:::||||||||||||||||||
     933 GGGCCCCTATGAAGGCTATAACCCCACGGTGAACCCTACTGTGTCCAACG 982

     484 alPheSerThrAlaAlaPheArgPheGlyHisAlaThrIleHisProLeu 500
         ||||||||||||||||||||||||||||||||||||||:::|||||||||
     983 TCTTCTCTACTGCTGCCTTCCGCTTTGGCCATGCCACAGTCCATCCACTG 1032

     501 ValArgArgLeuAspAlaSerPheGlnGluHisProAspLeuProGlyLe 517
         ||||||||||||:::::::::||||||:::|||    :::||||||    ||
    1033 GTGAGACGACTAAACACGGACTTCCAGGACCACACAGAGCTCCCCAGGCT 1082

     517 uTrpLeuHisGlnAlaPhePheSerProTrpThrLeuLeuArgGlyGlyG 534
         |    ||||||:::::||||||    ||||||    |||::::::    ||||
    1083 GCAGCTGCACGACGTCTTCTTCAGGCCCTGGAGGCTTATCCAGGAAGGTG 1132

     534 lyLeuAspProLeuIleArgGlyLeuLeuAlaArgProAlaLysLeuGln 550
         |||||||||:::::|||||||||||||||||||||||||||||||||||
    1133 GTTTGGATCCAATAGTAAGAGGCCTCCTGGCAAGACCAGCCAAGCTGCAA 1182

     551 ValGlnAspGlnLeuMetAsnGluGluLeuThrGluArgLeuPheValLe 567
         ||||||:::|||||||||||||||||||||||||||||||||||||||||
    1183 GTACAGGAGCAACTGATGAATGAGGAACTGACCGAGAGGCTCTTTGTGTT 1232
```

```
 567 uSerAsnSerSerThrLeuAspLeuAlaSerIleAsnLeuGlnArgGlyA 584
     |||||||    :::|||||||||||||||||||:::||||||||||||||||
1233 GTCTAATGTGGGCACCTTGGATCTGGCATCACTGAACTTGCAGAGGGGCC 1282

 584 rgAspHisGlyLeuProGlyTyrAsnGluTrpArgGluPheCysGlyLeu 600
     ||||||||||||||||||||||||||||||||||||||||||||||||||
1283 GGGACCACGGCTTACCAGGCTACAATGAGTGGAGAGAGTTCTGTGGCTTG 1332

 601 ProArgLeuGluThrProAlaAspLeuSerThrAlaIleAlaSerArgSe 617
        |||||||||||||||||:::|||:::    |||||||||:::|||||
1333 TCACGCCTGGAGACACCAGCTGAGCTGAACAAGGCCATTGCCAACAGAAG 1382

 617 rValAlaAspLysIleLeuAspLeuTyrLysHisProAspAsnIleAspV 634
     |::::::::||||||||:::::||||||||||||     ||||||||||||
1383 CATGGTCAACAAGATAATGGAGTTATACAAGCATGCTGACAACATTGATG 1432

 634 alTrpLeuGlyGlyLeuAlaGluAsnPheLeuProArgAlaArgThrGly 650
     |||||||||||||||||||||:::|||||||||      ||||||||||||
1433 TCTGGTTGGGAGGCCTGGCTGAAAAGTTCTTACCGGGAGCCCGCACCGGT 1482

 651 ProLeuPheAlaCysLeuIleGlyLysGlnMetLysAlaLeuArgAspGl 667
     ||||||||||||||:::|||||||||||||||||||||||||||||||||
1483 CCTCTGTTTGCATGTATCATTGGGAAGCAGATGAAGGCTCTGAGGGATGG 1532

 667 yAspTrpPheTrpTrpGluAsnSerHisValPheThrAspAlaGlnArgA 684
     ||||   ||||||||||||||||||||||||||||||||||||||||||:
1533 GGACAGGTTTTGGTGGGAGAACAGCCATGTCTTCACAGACGCTCAGAGGC 1582

 684 rgGluLeuGluLysHisSerLeuSerArgValIleCysAspAsnThrGly 700
     ::||||||||||||||||||||     |||||||||||||||||||||||
1583 AGGAACTGGAAAAGCATTCACTACCTCGGGTCATCTGTGACAACACCGGC 1632

 701 LeuThrArgValProMetAspAlaPheGlnValGlyLysPheProGluAs 717
     ||||||||||||||||||||||||:::||||||||||||||||||:::||
1633 CTCACCAGAGTACCTGTGGATGCCTTCCGTATTGGAAAGTTCCCCCAGGA 1682

 717 pPheGluSerCysAspSerIleThrGlyMetAsnLeuGluAlaTrpArgG 734
     ||||||||||||::::::|||    :::|||:::|||:::    |||||||
1683 CTTTGAATCCTGTGAGGAAATCCCTAGCATGGACCTCAGACTGTGGAGGG 1732

 734 luThrPheProGlnAspAspLysCysGlyPheProGluSerValGluAsn 750
     |||||||||||||||||||||||||    |||||||||:::|||:::|||
1733 AGACCTTCCCACAAGACGACAAGTGTGTCTTCCCAGAGAAGGTGGACAAT 1782

 751 GlyAspPheValHisCysGluGluSerGlyArgArgValLeuValTyrSe 767
     |||:::|||||||||||||||||||||||||:::     ||||||||||||||
1783 GGGAACTTTGTGCACTGTGAAGAATCTGGGAAGCTGGTACTGGTGTATTC 1832

 767 rCysArgHisGlyTyrGluLeuGlnGlyArgGluGlnLeuThrCysThrG 784
     ||||    ||||||||:::|||||||||:::|||||||:::||||||||||
1833 CTGTTTCCATGGATACAAGCTGCAAGGCCAGGAGCAGGTCACATGTACCC 1882

 784 lnGluGlyTrpAspPheGlnProProLeuCysLysAspValAsnGluCys 800
     ||:::||||||||||    :::||||||||||||||||||||||||||||
1883 AGAATGGATGGGACTCAGAGCCTCCTGTCTGTAAAGATGTTAATGAGTGT 1932

 801 AlaAspGlyAlaHisProProCysHisAlaSerAlaArgCysArgAsnTh 817
     ||||||    :::||||||||||||||||:::||||||:::|||:::|||||
1933 GCAGATCTGACACACCCACCTTGCCACTCCTCCGCAAAGTGCAAGAACAC 1982

 817 rLysGlyGlyPheGlnCysLeuCysAlaAspProTyrGluLeuGlyAspA 834
     |||||||:::|||||||||:::|||:::|||||||||    |||||||:::|
1983 CAAGGGAAGCTTCCAGTGTGTGTGCACAGACCCCTACATGCTAGGTGAGG 2032

 834 spGlyArgThrCysValAspSerGlyArgLeuProArgValThr 848
     ||    :::|||||||:::|||||||||||||||||||||||::::::
2033 ATGAGAAGACCTGCATAGATTCTGGCAGGCTACCTCGGGCATCC 2076
```

seq_name: gb_in:AB022197

seq_documentation_block:

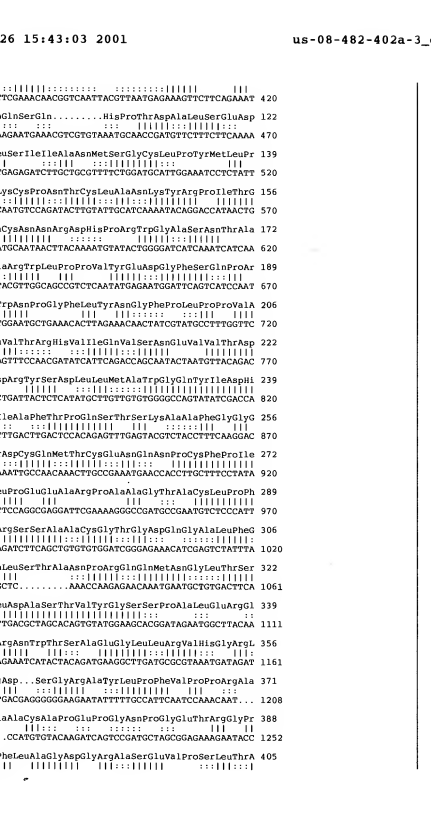
```
LOCUS       AB022197     3179 bp    mRNA            INV       02-APR-1999
DEFINITION  Halocynthia roretzi mRNA for homologue of mammlian thyroid
            peroxidase, complete cds.
ACCESSION   AB022197
VERSION     AB022197.1  GI:4587264
KEYWORDS    homologue of mammlian thyroid peroxidase.
SOURCE      Halocynthia roretzi endostyle tissue_lib:endostyle cDNA to mRNA,
            clone:HrTPO.
  ORGANISM  Halocynthia roretzi
            Eukaryota; Metazoa; Chordata; Urochordata; Ascidiacea;
            Stolidobranchia; Pyuridae; Halocynthia.
REFERENCE   1  (sites)
  AUTHORS   Ogasawara,M., Di Lauro,R. and Satoh,N.
  TITLE     Ascidian homologue of the mammalian thyroid peroxidase genes are
            expressed in the thyroid equivalent region of endostyle
  JOURNAL   Unpublished (1999)
REFERENCE   2  (bases 1 to 3179)
  AUTHORS   Ogasawara,M., Di Lauro,R. and Satoh,N.
  TITLE     Direct Submission
  JOURNAL   Submitted (10-JAN-1999) to the DDBJ/EMBL/GenBank databases. Michio
            Ogasawara, Kyoto University, Department of Zoology, Graduate School
            of Science; Kitashirakawaoiwake-cho, Sakyo-ku, Kyoto, Kyoto
            606-8052, Japan (E-mail:ogasawara@ascidian.zool.kyoto-u.ac.jp,
            Tel:81-75-753-4095, Fax:81-75-705-1113)
FEATURES             Location/Qualifiers
     source          1..3179
                     /organism="Halocynthia roretzi"
                     /db_xref="taxon:7729"
                     /clone="HrTPO"
                     /tissue_lib="endostyle"
                     /tissue_type="endostyle"
     CDS             66..2822
                     /codon_start=1
                     /product="homologue of mammlian thyroid peroxidase"
                     /protein_id="BAA76689.1"
                     /db_xref="GI:4587265"
                     /translation="MSSNSLPSLSLGMLLVLFFVLSKIRNARCYEDHFQDTNRMQAEG
                     DEFATKCMYEAMLNVNRAIRRTDIHWRKDTVWDSPMDLFQFFKRPPNEDAERMARASD
                     IFETTVNYVNEKVLQKYRRMKRRVNATDVLSSKTLRDLAAFSGCIGNPLLKQCPDTCI
                     ASKYRTITGQCNNLQNVYWGSSNHQLVRWQPSQYENGFSHPIGWNAETLRNNYRMPLV
                     RKVSNDIIQTSNTNVTDDTDYSHMLVVWGQYIDHDFDLTPQSLSTSTFQGLTNCQQTC
                     RNEPPCFPILLPGEDSKRADAECLPFFRSSAVCGSGETSSLFNELKPREQMNAVTSFL
                     DASTVYGSTDRMAYNLRNHTTDEGLMRVNDRFYDEGGRIFLPFNPNNPCVQDQSDASG
                     ERIPCFTAGDPRVSEHLTLSAIHTLWVRAHNRIARELKRINPHWYGETIYQEARKIVG
                     SLHQIVHYKEYVPKIIGMTGMNLLGEYSEYNPSVNPTISNVFATAAFRFGHVTIAPIF
                     RRLDGNFNEHPTHGNIFLHEAFFSPWRIIRQGGLDPIFRGLIGRPAKLITGTQIMHEE
                     LREKLFQLQNKVALDLASLNLQRGRDHAIPLYSYWREFCNLTRVETFDELASEISDAS
                     VELNWQNYTGHPGNLDLWLAGLVEDLVPGSRVGPTFLCLLTKQFQYLRDGDRFFYERV
                     HTDEQIEELEKIRLANVLCYNSGLETVQRDVFSLAQYPDDFVRCSDLDPLNLEPWREE
                     PEVGSCGRPQNIEYGDWQRCNDLVSYRCKMGFYLDGQEELTCMDNGAFNAEPPMCVDV
                     NECDQELKCNCEDICMNTVGSCRCMCSDGKILNEDGRTCSESPVTVVAPVGTNVAAIV
                     TGVVLGVALLVLVVAVTYGVHKYTLLMQVATQAGTSSVNTVKMGISNSGFDSSSIDKT
                     NMMH"
     polyA_site      3179
                     /note="19 a nucleotides"
BASE COUNT      974 a    617 c    700 g    888 t
ORIGIN
```

alignment_scores:

```
          Quality: 1927.50          Length:    771
            Ratio:  3.323             Gaps:     10
 Percent Similarity:  75.227   Percent Identity: 48.119
```

alignment_block:

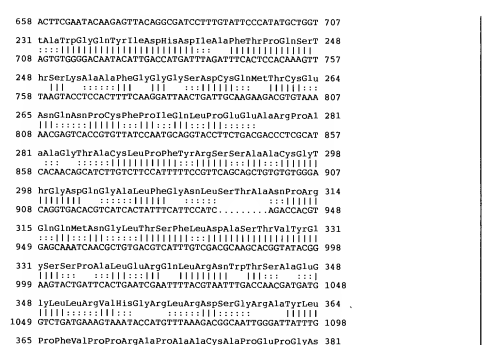
```
 US-08-482-402A-3_COPY_1_848 x AB022197 ..

 Align seg 1/1  to: AB022197  from: 1  to: 3179

      76 PheSerLysLeuProGluProThrSerGlyValIleAlaArgAlaAlaGl 92
         |||:::::::   |||:::        :::       :::|||||||||:::::
     321 TTCAAACGACCCCCGAACGAAGATGCTGAGAGGATGGCGAGAGCTTCAGA 370

      92 uIleMetGluThrSerIleGlnAlaMetLysArgLysValAsnLeuLysT 109
```



```
 705 roMetAspAlaPheGlnValGlyLysPheProGluAspPheGluSerCys 721
         |||:::|||:::::::::::::::|||:::||||||       |||
2194 AACGTGACGTATTCTCACTTGCGCAGTATCCTGATGATTTTGTCCGATGC 2243

 722 AspSerIleThrGlyMetAsnLeuGluAlaTrpArgGluThrPheProGl 738
     ::::::::      :::|||||||||  |||||||||     |||::
2244 TCTGATTTGGACCCGTTGAACTTGGAGCCATGGAGAGAAGAA...CCAGA 2290

 738 nAspAspLysCysGlyPheProGluSerValGluAsnGlyAspPheValH 755
     :      :::||||||  |||:::::::::|||  |||||||:::    :
2291 AGTTGGTTCATGTGGAAGGCCGCAAAACATTGAGTATGGCGACTGGCAAA 2340

 755 isCysGluGluSerGlyArgArgValLeuValTyrSerCysArgHisGly 771
     ::|||::::::              ::::::   |||   |||:::   |||
2341 GGTGCAATGAC............CTCGTCAGTTATAGATGCAAAATGGGT 2378

 772 TyrGluLeuGlnGlyArgGluGlnLeuThrCysThrGlnGluGly...Tr 787
     :::    |||:::|||:::|||:::|||||||||   :::::::|||   ::
2379 TTTTACTTGGATGGACAAGAGGAGTTAACATGTATGGATAATGGTGCTTT 2428

 787 pAspPheGlnProProLeuCysLysAspValAsnGluCysAlaAspGlyA 804
     ::::    :::|||||||:::|||    ||||||||||||||
2429 TAATGCAGAACCACCAATGTGTGTTGATGTGAACGAATGT......GACC 2472

 804 laHisProProCysHisAlaSerAlaArgCysArgAsnThrLysGlyGly 820
       :::       |||:::::::::       |||   ||||||   |||:::
2473 AGGAATTGAAATGTAATTGTGAGGATATATGCATGAACACTGTTGGTAGC 2522

 821 PheGlnCysLeuCysAlaAspProTyrGluLeuGlyAspAspGlyArgTh 837
       :::|||:::|||:::|||            |||:::::::|||||||||||
2523 TGTCGATGTATGTGTTCTGATGGCAAAATATTAAACGAGGATGGAAGAAC 2572

 837 rCysValAspSer 841
     ||||    :::|||
2573 TTGTTCTGAGTCC 2585
```

seq_name: gb_in:AB022196

```
seq_documentation_block:
LOCUS       AB022196     2915 bp    mRNA             INV       02-APR-1999
DEFINITION  Ciona intestinalis mRNA for homologue of mammlian thyroid
            peroxidase, complete cds.
ACCESSION   AB022196
VERSION     AB022196.1  GI:4587262
KEYWORDS    homologue of mammlian thyroid peroxidase.
SOURCE      Ciona intestinalis endostyle tissue_lib:endostyle cDNA to mRNA,
            clone:CiTPO.
  ORGANISM  Ciona intestinalis
            Eukaryota; Metazoa; Chordata; Urochordata; Ascidiacea;
            Phlebobranchia; Cionidae; Ciona.
REFERENCE   1  (sites)
  AUTHORS   Ogasawara,M., Di Lauro,R. and Satoh,N.
  TITLE     Ascidian homologue of the mammalian thyroid peroxidase genes are
            expressed in the thyroid equivalent region of endostyle
  JOURNAL   Unpublished (1999)
REFERENCE   2  (bases 1 to 2915)
  AUTHORS   Ogasawara,M., Di Lauro,R. and Satoh,N.
  TITLE     Direct Submission
  JOURNAL   Submitted (10-JAN-1999) to the DDBJ/EMBL/GenBank databases. Michio
            Ogasawara, Kyoto University, Department of Zoology, Graduate School
            of Science; Kitashirakawaoiwake-cho, Sakyo-ku, Kyoto, Kyoto
            606-8052, Japan (E-mail:ogasawara@ascidian.zool.kyoto-u.ac.jp,
            Tel:81-75-753-4095, Fax:81-75-705-1113)
FEATURES             Location/Qualifiers
     source          1..2915
                     /organism="Ciona intestinalis"
                     /db_xref="taxon:7719"
                     /clone="CiTPO"
                     /tissue_lib="endostyle"
                     /tissue_type="endostyle"
     CDS             19..2748
                     /codon_start=1
                     /product="homologue of mammlian thyroid peroxidase"
                     /protein_id="BAA76688.1"
                     /db_xref="GI:4587263"
                     /translation="MWTKCIAVVLFCKLVFADEQYVFDETGYEGDEFAREALREAFQQ
                     VRRAIFRTDKSWSDEKVWRNPMTLFQYFKRPPSEESSDMAKAGDIYHCAFDIVHDKVS
                     NRLSRFKRRVNVTDVLSSATIAELAKFSGCRANGVSPTCPDTCIDNMYRTHTGECNNK
                     VNRYWGASNNPFVRWRPAQYENEFSTPIGWDSQRSYNGYQLPVVRKVSNDIMRTSNTR
                     VTGDPLYSHMLVVWGQYIDHDLDFTPQSLSTSTFQGLTDCKKTCKNESPCYPMQVPSD
                     DPRITTASCLPFFRSAAVCGTGDTSSLFHSIRPREQINAVTSFVDASTVYGSTDSLNR
                     ILRNLTNDDGLMKVNTMFKDGNWDYLPFDPNNPCVQDPFDASGVNIPCFHAGDGRVSE
                     HLTLSAIHTVWVREHNRIARMLKSMNPHWSGEIIYQEARKIVGAYHQIVHWKEYVPKI
                     IGPAGLRMMGNYTGYRENENPTVSNVFATAAFRFGHATISPQFRRLDENFNNHPQFPT
                     ILLHEAFFSPWRMIREGGMDPILRGLIGRPAKLIKADEMMHEELRDKLFALQNQVALD
                     LASLNLQRGRDHALPLYNDWREECGLARANNFSDLAGEIKDKAIRDKLEALYGHPGNI
                     DLWLAGLSEDLMDGSRGGPVFTCLLARQFKFLRNGDRFYYENPNVFTPNQVTALNRLS
                     FARVLCDNSGLTRVQPDLFMLRDTSQFVDCANIPNLDLSQWREDPAVGTCGTPPSISN
                     GWWKKCGGGVSYRCMSGYQLAGPNEVQCMNSAFTAIPSCVDINECDAQNGGCSDRCMN
                     TAGSYICVCNDDRMLGSDGKTCMATTVQVTPSPAVMGVTATTPANVTAIVVGSILGAA
                     IFLAFCVICYLVYKNTKLHTKSKYHNTGKFSLSDPSAYDNPSAVMDNDSQTKM"
     polyA_site      2915
                     /note="20 a nucleotides"
BASE COUNT      833 a    610 c    687 g    785 t
ORIGIN


alignment_scores:
            Quality:  1879.00          Length:    775
              Ratio:  3.302              Gaps:     11
 Percent Similarity:  73.419   Percent Identity: 46.323

alignment_block:
 US-08-482-402A-3_COPY_1_848 x AB022196  ..

 Align seg 1/1  to: AB022196  from: 1  to: 2915

  76 PheSerLysLeuProGluProThrSerGlyValIleAlaArgAlaAlaGl 92
     |||::::::   |||:::      |||:::   :::|||:::|||:::::
 232 TTTAAACGACCGCCCAGCGAAGAATCGAGCGATATGGCGAAAGCTGGGGA 281

  92 uIle..............................MetGluThrSerI 98
     :|||                              :::::::::   :
 282 TATTTATCATTGTGCGTTTGATATTGTACATGACAAGGTGTCGAACCGAC 331

  98 leGlnAlaMetLysArgLysValAsnLeuLysThrGlnGlnSerGlnHis 114
     :::::   :::|||||||:::|||||||:::
 332 TTTCAAGGTTTAAACGCCGGGTAAACGTC..................... 360

 115 ProThrAspAlaLeuSerGluAspLeuLeuSerIleIleAlaAsnMetSe 131
        ||||||:::||||||:::        ::::::   :::|||:::::::||
 361 ...ACTGATGTGCTGTCAAGCGCAACGATCGCAGAGTTGGCCAAATTTTC 407

 131 rGlyCysLeuProTyrMetLeuProProLysCysProAsnThrCysLeuA 148
     |||||||           :::   |||   ||||||:::||||||:::
 408 TGGTTGTCGTGCAAATGGTGTGAGTCCAACATGTCCGGATACTTGTATCG 457

 148 laAsnLysTyrArgProIleThrGlyAlaCysAsnAsnArgAspHisPro 164
       |||   ||||||      ||||||   |||||||||:::   :::
 458 ACAATATGTACAGAACCCACACCGGTGAATGCAACAACAAAGTTAATCGA 507

 165 ArgTrpGlyAlaSerAsnThrAlaLeuAlaArgTrpLeuProProValTy 181
        |||||||||||||||:::   ::::::||||||   |||       ||
 508 TACTGGGGTGCTTCGAATAATCCGTTCGTTAGATGGCGCCCTGCACAATA 557

 181 rGluAspGlyPheSerGlnProArgGlyTrpAsnProGlyPheLeuTyrA 198
     ||||:::   ||||||   |||   |||||||:::           ||||
 558 CGAAAATGAATTCAGCACTCCAATAGGTTGGGATAGCCAACGTTCGTATA 607

 198 snGlyPheProLeuProProValArgGluValThrArgHisValIleGln 214
     ||||||:::   ||||||   ||||||:::|||::::::   :::::::::
 608 ATGGATACCAATTACCTGTTGTGCGAAAAGTTTCAAACGATATAATGAGA 657

 215 ValSerAsnGluValValThrAspAspAspArgTyrSerAspLeuLeuMe 231
     :::||||||       ||||||   |||       ||||||   :::|||::
```


658 ACTTCGAATACAAGAGTTACAGGCGATCCTTTGTATTCCCATATGCTGGT 707
231 tAlaTrpGlyGlnTyrIleAspHisAspIleAlaPheThrProGlnSerT 248
707 AGTGTGGGGACAATACATTGACCATGATTTAGATTTCACTCCACAAAGTT 757
248 hrSerLysAlaAlaPheGlyGlySerAspCysGlnMetThrCysGlu 264
758 TAAGTACCTCCACTTTTCAAGGATTAACTGATTGCAAGAAGACGTGTAAA 807
265 AsnGlnAsnProCysPhePheIleGlnLeuProGluGlyAlaArgProAl 281
808 AACGAGTCACCGTGTTATCCAATGCAGGTACCTTCTGACGACCCTCGCAT 857
281 aAlaGlyThrAlaCysLeuProPheTyrArgSerSerAlaAlaCysGlyT 298
858 CACAACAGCATCTTGTCTTCCATTTTCCGTTCAGCAGCTGTGTGTGGGA 907
298 hrGlyAspGlnGlyAlaLeuPheGlyAsnLeuSerThrAlaAsnProArg 314
907 CAGGTGACAGCTCATCACTATTTCATTCCATC.....AGACCACGT 948
315 GlnGlnMetAsnGlyLeuThrSerPheLeuAspAlaSerThrValTyrGl 331
949 GAGCAAATCAACGCTGTGACGTCATTTGTCGACGCAAGCACGGTATACGG 998
331 ySerSerProAlaLeuGluArgGlnLeuArgAsnTrpThrSerAlaGluG 348
999 AAGTACTGATTCACTGAATCGAATTTTACGTAATTTGACCAACGATGATG 1048
348 lyLeuLeuArgValHisGlyArgLeuArgAspSerGlyArgAlaTyrLeu 364
1049 GTCTGATGAAAGTAAATACCATGTTTAAAGACGGCAATTGGGATTATTTG 1098
365 ProPheValProProAsrgAlaProAlaAlaCysAlaProGluProGlyAs 381
1099 CCATTCGATCCGAACAAC.....CCGTGTGTACAAGATCCTTTCGA 1139
381 nProGlyGluThrArgGlyProCysPheLeuAlaGlyAspGlyArgAlaS 398
1140 TGCGAGCGGCGTCAACATTCCTTGTTTCCACGCAGGAGATGGAAGAGTAA 1189
398 erGluValProSerLeuThrAlaLeuHisThrLeuTrpLeuArgGluHis 414
1190 GCGAGCATTTAACATTAAGTGCAATACACACAGTATGGGTGCGAGAACAC 1239
415 AsnArgLeuAlaAlaAlaLeuLysAlaLeuAsnAlaHisTrpSerAlaAs 431
1240 AACAGGATTGCTCGTATGTTAAAGTCCATGAACCCACACTGGTCGGGTGA 1289
431 pAlaValTyrGlnGluAlaArgLysValValGlyAlaLeuHisGlnIleI 448
1290 AATTATTTACCAGGAAGCAAGGAAGATAGTTGGAGCTTATCACCAGATCG 1339
448 leThrLeuArgAspTyrIleProArgIleLeuGlyProGluAlaPheGln 464
1340 TGCATTGGAAGGATACGTGCCAAAGATAATCGGCCCAGCTGGTTTGCGC 1389
465 GlnTyrValGlyProTyrGluGlyTyrAspSerThrAlaAsnProThrVa 481
1390 ATG...ATGGGCAACTATACAGGTTATAGAGAGAATGAAAACCCAACAGT 1436
481 lSerAsnValPheSerThrAlaAlaPheArgPheGlyHisAlaThrIleH 498
1437 TTCAACGTTTTCGCCACAGCTGCGTTTCGTTTCGGCCATGCGACTATCT 1486
498 isProLeuValArgArgLeuAspAlaSerPheGlnGluHisProAspLeu 514
1487 CGCCCCAGTTCAGGAGACTGGACGAAAACTTTAACAATCACCCGCAGTTC 1536
515 ProGlyLeuTrpLeuHisGlnAlaPhePheSerProTrpThrLeuLeuAr 531
1537 CCGACCATTCTCTTTGCATGAAGCTTTTTCTCCCCATGGCGAATGATAAG 1586


531 gGlyGlyGlyLeuAspProLeuIleArgGlyLeuAlaArgProAlaL 548
1587 GGAGGGGGAATGGATCCTATATTGAGAGGTTTGATCGGCCGACCGGCCA 1636
548 ysLeuGlnValGlnAspGlnLeuMetAsnGluGluLeuThrGluArgLeu 564
1637 AACTTATCAAAGCTGACGAAATGATGCACGAAGAGCTTAGGGATAAACTC 1686
565 PheValLeuSerAsnSerSerThrLeuAspLeuAlaSerIleAsnLeuGl 581
1687 TTGCTTTACAAAACCAGGTGCATTGGATTTGCCTCACTGAATCTACA 1736
581 nArgGlyArgAspHisGlyLeuProGlyTyrAsnGluTrpArgGluPheC 598
1737 GAGAGGTCGTGACCACGCGTTACCATTATACAATGATTGGCGTGAGGAGT 1786
598 ysGlyLeuProArgLeuGluThrProAlaAspLeuSerThrAlaIleAla 614
1787 GTGGTTAGCAAGGGCGAACAACTTCAGTGACTTGGCTGGTGAAATAAAA 1836
615 SerArgSerValAlaAspLysIleLeuAspLeuTyrLysHisProAspAs 631
1837 GACAAACAATTCGTGATAAACTAGAAGCTCTATACGGTCATCCAGGAAA 1886
631 nIleAspValTrpLeuGlyGlyLeuAlaGluAsnPheLeuProArgAlaA 648
1887 CATAGATTTATGGTTAGCTGGTCTTTCCGAAGATTTAATGGACGGTTCGC 1936
648 rgThrGlyProLeuPheAlaCysLeuIleGlyLysGlnMetLysAlaLeu 664
1937 GTGGTGGACCAGTCTTCACTTGCCTGCTTGCTCGTCAGTTTAAATTCCTT 1986
665 ArgAspGlyAspTrpPheTrpTrpGluAsnSerHisValPheThrAspAl 681
1987 CGAAACGGAGACAGATTTTACTACGAGAATCCAAATGTTTCACACCGAA 2036
681 aGlnArgArgGluLeuGluLysHisSerLeuSerArgValIleCysAspA 698
2037 TCAGGTCACAGCGTTAAACAGATTATCCTTCGCCCGTGTTGTTGTGATA 2086
698 snThrGlyLeuThrArgValProMetAspAlaPheGlnValGlyLysPhe 714
2087 ACTCTGGTTTAACGCGAGTCCAACCCGACTTGTTCATGCTA...CGTGAT 2133
715 ProGluAspPheGluSerCysAspSerIleThrGlyMetAsnLeuGluAl 731
2134 ACATCTCAGTTTGTTGATGTGTGCAAACATTCCTAACCTGGATCTTTCACA 2183
731 aTrpArgGluThrPheProGlnAspAspLysCysGlyPhePheGluSerV 748
2184 GTGGCTGAAGAC...CCAGCAGTTGGTACATGCGGTACACCACCTAGTA 2230
748 alGluAsnGlyAspPheValHisCysGluGluSerGlyArgArgValLeu 764
2231 TAAGTAACGGTTGGTGGAAGAAGTGTGGTGGGGGT.....GTT 2268
765 ValTyrSerCysArgHisGlyTyrGluLeuGlnGlyArgGluGlnLeuTh 781
2269 AGTTATCGTTGTATGTCTGGATTACCAACTTGCTGGACCAAATGAGGTGCA 2318
781 rCysThrGlnGluGlyTrpAspPheGlnProProLeuCysLysAspValA 798
2319 ATGTATGAACTCAGCTTTTACC...GCCATACCATCTTGTGTTGACATAA 2365
798 snGluCysAlaAspGlyAlaHisProProCysHisAlaSerAlaArgCys 814
2366 ACGAGTGC...GATGCACAGAACGGTGGGTGC.....AGTGACAGGTGT 2406
815 ArgAsnThrLysGlyGlyPheGlnCysLeuCysAlaAspProYrYrGluLe 831
2407 ATGAACACCGCTGGTAGTTACATTTGTGTTTGCAATGACGATCGTATGCT 2456

```
    831 uGlyAspAspGlyArgThrCysVal 839
        ||||:::||||||:::||||||:::
   2457 TGGTAGTGATGGTAAAACTTGCATG 2481

seq_name: gb_ov:AF349034

seq_documentation_block:
LOCUS       AF349034     3846 bp    mRNA            VRT       02-AUG-2001
DEFINITION  Danio rerio myeloperoxidase mRNA, complete cds.
ACCESSION   AF349034
VERSION     AF349034.1  GI:15077546
KEYWORDS    .
SOURCE      zebrafish.
  ORGANISM  Danio rerio
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Actinopterygii; Neopterygii; Teleostei; Euteleostei; Ostariophysi;
            Cypriniformes; Cyprinidae; Rasborinae; Danio.
REFERENCE   1  (bases 1 to 3846)
  AUTHORS   Langenau,D.M., Bennett,C.M., Deng,M., Kanki,J.P., Rhodes,J.,
            Liu,T.X., Delahaye-Brown,A. and Look,A.T.
  TITLE     Genetic Analysis of Myelopoiesis in the Zebrafish
  JOURNAL   Unpublished
REFERENCE   2  (bases 1 to 3846)
  AUTHORS   Langenau,D.M., Bennett,C.M., Deng,M., Kanki,J.P., Rhodes,J.,
            Liu,T.X., Delahaye-Brown,A. and Look,A.T.
  TITLE     Direct Submission
  JOURNAL   Submitted (13-FEB-2001) Pediatric Oncology, Dana-Farber Cancer
            Institute, 44 Binney Street, Boston, MA 02115, USA
FEATURES             Location/Qualifiers
     source          1..3846
                     /organism="Danio rerio"
                     /db_xref="taxon:7955"
     CDS             73..2361
                     /codon_start=1
                     /product="myeloperoxidase"
                     /protein_id="AAK83239.1"
                     /db_xref="GI:15077547"
                     /translation="MNLPTFLFVVGCCCALSVGAEEESPGRPFILDSIEEAKKIVDAA
                     YKYSRDESLARVRKDVIKPSDKLRLLKQPARKTREAVRAADYMAQTLRLISEKAHHVH
                     KRSINATNLLTQDELQTITRMTGCEAQTRPPSCRTTPLINKYRTASGVCNNRKNPLLG
                     ASNTPFARWLPAQYEDAVSQPKGWDPNKLYNGAALPMVRLVSNRILATADADIESDHD
                     FTFMLTIFGQWVDHDLTFTPFSPSIRSFSNGIDCENSCERSEPCFPISAPPGDQRLRP
                     NTCLPVFRSAPTCGSGHTAYMFGEVPNVREQINTLTAYLDAGQVYGSEDGLAKELRDL
                     TNDGGLLRVNNRFKDNGRELLPFTSVNTNLCATRQKILNDSTLTEVPCFIAGDARVDE
                     NPALNSLHTLFVREHNRLARALHVLNPTWSSETLYQEARKIVGAFNQILVIKEYLPLI
                     VGPDAYNRHLGPYPGYNENVDPTIANVFATAAFRFAHLTIQPFIFRLDENYKNHPQFP
                     SVPLYEAFFSPWRVIFEGGIDPVLRGLIGRPAKLNTQDHMLVNALRERLFAFTSHIAL
                     DLASLNMQRGRDHAIPGYNAWRRFCGLSAPQNEQELAVVMNNTKLARKLIELYGTPEN
                     IDVWLGGVAEPFAPGGRVGSLFACLISRQFQKIRDGDRLWFESNGVFTTKQKTALASV
                     SMARIICDNTGILKVPRDPFRFRSPASFVNCGNIPAFDLEPWKETGDGGINISDIGNE
                     IEIPSQDLPDPGDLQENVVHRV"
BASE COUNT     1062 a    850 c    826 g   1108 t
ORIGIN


alignment_scores:
            Quality: 1670.50             Length:    855
              Ratio:  2.808                Gaps:     15
 Percent Similarity:  69.591    Percent Identity: 40.234

alignment_block:
 US-08-482-402A-3_COPY_1_848 x AF349034   ..

 Align seg 1/1  to: AF349034  from: 1  to: 3846

     32 GlyLysProGluGluSerArg............ValSerSerValLeuGl 44
        |||      |||||||||               :::   :::   :::||
    127 GGGGCAGAAGAAGAAAGTCCTGGAAGACCTTTTATTCTTGATTCAATTGA 176

     44 uGluSerLysArgLeuValAspThrAlaMetTyrAlaThrMetGlnArgA 61
        ||||:::|||::::::||||||:::|||            :::  ::::::
    177 GGAAGCAAAGAAAATAGTGGACGCTGCTTACAAGTATTCTCGAGATGAGA 226

     61 snLeuLys......LysArgGlyIleLeuSerGlyAlaGlnLeuLeuSer 75
        ::|||          ::::::  :::::::::  ::::::   |||
    227 GTTTAGCAAGGGTGCGCAAAGATGTCATTAAGCCTTCTGACAAGCTACGT 276

     76 PheSerLysLeuProGluProThrSerGlyValIleAlaArgAlaAlaGl 92
        :::   |||   |||            :::         :::||||||||||::
    277 CTGCTGAAACAGCCAGCTCGAAAGACACGGGAGGCCGTAAGAGCTGCAGA 326

     92 uIleMetGluThrSerIleGlnAlaMetLysArgLysValAsnLeuLysT 109
        :   |||       :::::::::  :::::::::|||::::::
    327 CTACATGGCACAAACGCTGAGACTGATCAGTGAGAAAGCCCATCATGTAC 376

    109 hrGlnGlnSerGlnHisProThrAspAlaLeuSerGluAspLeuLeuSer 125
         ::::::|||   :::    |||:::   |||::::::|||   |||:::
    377 ACAAGAGGTCCATCAATGCCACAAACCTGCTCACTCAAGACGAGCTTCAA 426

    126 IleIleAlaAsnMetSerGlyCysLeuProTyrMetLeuProProLysCy 142
          |||::::::|||:::||||||                ||||||:::||
    427 ACCATTACACGTATGACAGGCTGTGAAGCTCAAACCCGGCCACCATCTTG 476

    142 sProAsnThrCysLeuAlaAsnLysTyrArgProIleThrGlyAlaCysA 159
        |    :::|||  |||    ||||||||||||      :::|||:::||||
    477 TAGGACCACACCCCTCATAAACAAGTATCGAACTGCCAGCGGTGTCTGCA 526

    159 snAsnArgAspHisProArgTrpGlyAlaSerAsnThrAlaLeuAlaArg 175
        ||||||||     :::|||    |||||||||||||||    :::||||||
    527 ACAATCGGAAGAACCCACTTCTTGGTGCATCCAACACTCCATTTGCCCGC 576

    176 TrpLeuProProValTyrGluAspGlyPheSerGlnProArgGlyTrpAs 192
        |||||||||       ||||||||||:::   |||||||||:::|||||||::
    577 TGGTTGCCTGCTCAATATGAGGACGCCGTTTCTCAGCCAAAAGGATGGGA 626

    192 nProGlyPheLeuTyrAsnGlyPheProLeuProProValArgGluValT 209
        :|||:::   ||||||||||||        ||||||   ||||||    |||:
    627 TCCCAACAAGTTGTACAATGGTGCAGCGCTGCCCATGGTTAGACTGGTTT 676

    209 hrArgHisValIleGlnValSerAsnGluValValThrAspAspAspArg 225
        :::::::::::::::  :::::::::          :::    :::|||
    677 CCAATCGCATTCTAGCCACTGCAGATGCTGACATAGAGAGTGACCATGAC 726

    226 TyrSerAspLeuLeuMetAlaTrpGlyGlnTyrIleAspHisAspIleAl 242
        ::::::   :::|||       :::||||||:::::::||||||||||:::::
    727 TTTACCTTCATGCTCACCATCTTTGGGCAATGGGTTGACCACGATCTGAC 776

    242 aPheThrProGlnSerThrSerLysAlaAlaPheGlyGlyGlySerAspC 259
        :|||||||||    |||    |||         :::|||::::::|||    ||||
    777 TTTTACACCTTTCTCACCAAGTATTAGGTCCTTTAGTAATGGCATTGACT 826

    259 ysGlnMetThrCysGluAsnGlnAsnProCysPheProIleGlnLeuPro 275
        ||:::    :::||||||::::::::::||||||||||||||||:::    |||
    827 GTGAAAACAGTTGTGAGCGCTCTGAACCCTGCTTCCCAATTTCGGCCCCT 876

    276 GluGluAlaArgProAlaAlaGlyThrAlaCysLeuProPheTyrArgSe 292
               :::                :::::||||||||||     :::|||||
    877 CCAGGAGATCAGCGTCTGCGCCCTAACACCTGTCTCCCTGTCTTCCGGTC 926

    292 rSerAlaAlaCysGlyThrGlyAspGlnGlyAlaLeuPheGlyAsnLeuS 309
        |:::    :::||||||:::|||          :::       :::|||||||::::::
    927 AGCACCGACTTGTGGTTCTGGGCACACTGCCTACATGTTTGGTGAAGTC. 975

    309 erThrAlaAsnProArgGlnGlnMetAsnGlyLeuThrSerPheLeuAsp 325
                |||    |||:::|||:::|||    ||||||::::::||||||
    976 .....CCAAATGTTCGGGAACAGATCAATACTCTAACAGCTTACCTAGAT 1020

    326 AlaSerThrValTyrGlySerSerProAlaLeuGluArgGlnLeuArgAs 342
        |||:::    |||||||||||||:::    :::|||    ::::::||||||::
   1021 GCTGGACAAGTTTACGGTTCAGAGGATGGACTAGCAAAGGAACTGCGGGA 1070

    342 nTrpThrSerAlaGluGlyLeuLeuArgValHisGlyArgLeuArgAspS 359
        :    |||:::          |||||||||||||||::::::|||::::::|||:
   1071 CCTTACTAATGATGGTGGTCTTTTACGTGTCAATAATCGTTTTAAAGACA 1120

    359 erGlyArgAlaTyrLeuProPheValProProArgAlaProAlaAlaCys 375
```

```
     ::||||||      |||||||||:::      ::::::         |||
1121 ATGGACGAGAGCTCCTGCCCTTTACTAGTGTTAACACC...AACTTGTGC 1167

 376 AlaProGluPro.........GlyAsnProGlyGluThrArgGlyProCy 389
     |||  .:::            ::::::          |||:::  |||||
1168 GCCACTCGTCAGAAAATCCTCAACGACAGCACTCTGACAGAGGTTCCTTG 1217

 389 sPheLeuAlaGlyAspGlyArgAlaSerGluValProSerLeuThrAlaL 406
     ||||:::|||||||||:::|||::::::|||   |||:::|||::::::|
1218 CTTCATTGCGGGTGATGCTCGTGTCGATGAGAATCCTGCTCTGAACTCCT 1267

 406 euHisThrLeuTrpLeuArgGluHisAsnArgLeuAlaAlaAlaLeuLys 422
     ||||||||||::::::|||||||||||||||||||    ||||||
1268 TACACACACTTTTTGTAAGGGAGCACAATAGATTGGCTCGTGCCCTTCAT 1317

 423 AlaLeuAsnAlaHisTrpSerAlaAspAlaValTyrGlnGluAlaArgLy 439
     :::||||||       |||||::::::::::::|||||||||||||||
1318 GTGCTCAACCCTACATGGAGTAGTGAAACTCTTTATCAAGAAGCAAGAAA 1367

 439 sValValGlyAlaLeuHisGlnIleIleThrLeuArgAspTyrIleProA 456
     |:::|||||||||:::::|||||:::::::::::::::|||:::|||
1368 AATTGTTGGTGCATTCAACCAGATCTTGGTGATCAAGGAATACTTGCCAC 1417

 456 rgIleLeuGlyProGluAlaPheGlnGlnTyrValGlyProTyrGluGly 472
       |||:::||||||:::|||::::::::::::||||||||    |||
1418 TAATTGTAGGCCCTGATGCTTACAATAGACACCTTGGACCATATCCAGGT 1467

 473 TyrAspSerThrAlaAsnProThrValSerAsnValPheSerThrAlaAl 489
     |||:::::::::::::::||||||::::::|||||||||:::||||||||
1468 TACAATGAGAATGTGGACCCTACCATTGCGAACGTCTTTGCCACTGCTGC 1517

 489 aPheArgPheGlyHisAlaThrIleHisProLeuValArgArgLeuAspA 506
     |||||||||||:::|||    ||||||:::|||::::::    |||||||||
1518 TTTCCGTTTTGCCCACTTGACTATCCAGCCCTTCATTTTTCGTCTTGATG 1567

 506 laSerPheGlnGluHisProAspLeuProGlyLeuTrpLeuHisGlnAla 522
       ::::::::::::||||||::::::|||::::::    |||::::::|||
1568 AAAACTACAAGAACCATCCTCAGTTTCCCAGTGTCCCCCTCTATGAGGCT 1617

 523 PhePheSerProTrpThrLeuLeuArgGlyGlyGlyLeuAspProLeuIl 539
     |||||||||||||||    ::::::       ||||||:::|||||||::::
1618 TTCTTCTCTCCATGGAGGGTGATCTTTGAAGGGGGAATTGATCCTGTGCT 1667

 539 eArgGlyLeuLeuAlaArgProAlaLysLeuGlnValGlnAspGlnLeuM 556
     :||||||||::::::||||||||||||||||:::::::||||||:::::::
1668 CCGTGGATTGATTGGTCGTCCGGCAAAACTGAACACCCAGGACCACATGT 1717

 556 etAsnGluGluLeuThrGluArgLeuPheValLeuSerAsnSerSerThr 572
     ::    :::    |||    ||||||||||||:::::::::::          :::
1718 TGGTGAATGCTCTCAGAGAGAGGCTGTTTGCCTTCACATCCCACATAGCT 1767

 573 LeuAspLeuAlaSerIleAsnLeuGlnArgGlyArgAspHisGlyLeuPr 589
     |||||||||||||||:::|||:::||||||||||||||||||:::::::||
1768 TTGGATCTGGCTTCCCTCAACATGCAAAGAGGGCGTGACCATGCTATACC 1817

 589 oGlyTyrAsnGluTrpArgGluPheCysGlyLeuProArgLeuGluThrP 606
     ||||||||||     ||||||:::||||||||||||         ::::::
1818 AGGTTATAATGCATGGCGTCGGTTCTGTGGATTGTCTGCTCCTCAGAATG 1867

 606 roAlaAspLeuSerThrAlaIleAlaSerArgSerValAlaAspLysIle 622
          :::|||::::::::::::    :::    ::::::|||   |||:::
1868 AGCAAGAATTGGCTGTGGTGATGAACAACACTAAACTAGCCCGCAAGTTA 1917

 623 LeuAspLeuTyrLysHisProAspAsnIleAspValTrpLeuGlyGlyLe 639
     ::::::||||||      |||:::|||||||||||||||||||||||::
1918 ATTGAGCTTTATGGCACCCCTGAGAACATTGATGTTTGGTTAGGAGGTGT 1967

 639 uAlaGluAsnPheLeuProArgAlaArgThrGlyProLeuPheAlaCysL 656
     :||||||    |||    |||    :::|||:::|||    |||||||||||||
1968 TGCTGAGCCTTTTGCTCCTGGTGGTCGTGTCGGTTCTCTTTTGCCTGCC 2017

 656 euIleGlyLysGlnMetLysAlaLeuArgAspGlyAspTrpPheTrpTrp 672
     |||||::::::|||::::::    :::|||||||||||    :::|||:::

2018 TGATTTCAAGACAGTTCCAGAAAATCCGAGATGGCGATAGGTTGTGGTTT 2067

 673 GluAsnSerHisValPheThrAspAlaGlnArgArgGluLeuGluLysHi 689
     |||::::::   |||||||||       |||:::       |||    :::
2068 GAGAGCAATGGAGTGTTCACCACCAAACAGAAAACAGCTCTGGCTTCTGT 2117

 689 sSerLeuSerArgValIleCysAspAsnThrGlyLeuThrArgValProM 706
       ||||:::::|||:::||||||||||||||||||:::     :::||||||
2118 GTCAATGGCCCGCATAATCTGCGACAACACCGGAATCCTTAAAGTTCCCA 2167

 706 etAspAlaPheGlnValGlyLysPheProGluAspPheGluSerCysAsp 722
       |||    |||:::       :::    |||    :::|||    :::|||
2168 GAGATCCTTTTCGCTTC...CGGAGTCCTGCCAGTTTTGTTAACTGTGGG 2214

 723 SerIleThrGlyMetAsnLeuGluAlaTrpArgGluThr........... 735
     :::|||    :::::::::||||||    |||:::||||||
2215 AATATCCCAGCTTTTGATCTCGAGCCTTGGAAAGAGACTGGTGACGGTGG 2264

 736 .................................PheProGlnAspAspL 741
                                      :::|||:::::|||
2265 CATCAACATTTCAGACATTGGAAATGAGATAGAGATTCCATCCCAGGACC 2314

 741 ysCysGlyPheProGluSerValGluAsnGlyAspPheValHisCysGlu 757
                 :::       ||||||       ||||||
2315 TCCCGGACCCAGGGGACCTCCAGGAGAAC......GTGGTCCACAGGGTG 2358

 758 GluSerGlyArgArgValLeuValTyrSerCysArgHisGlyTyrGluLe 774
       :::|||::::::                          :::||
2359 TAGCAGGGCCACCAG...........................GACCT 2378

 774 uGlnGlyArgGluGlnLeuThrCysThrGlnGluGlyTrpAspPheGlnP 791
     ||||      :::::|||       |||:::       |||    :::
2379 CCAGGAATTCCAGGACCTCCCATTAATACAACAG......GACAACAGTC 2422

 791 roPro.LeuCysLysAspValAsnGluCysAlaAspGlyAlaHisProPr 807
       ||| |||                   |||    :::      ||||||
2423 TGCCTTCTT.................TGCCTCAGTCAACTCCATCCTTC 2454

 807 oCysHis.AlaSerAla................................ 812
       |||||| |||:::::
2455 CTGCCACTGCTAAAGTTGTGGTGTTCGGTCCGGTCCTTTACAATGGACAA 2504

 813 ...ArgCysArgAsnThrLysGlyGlyPheGlnCysLeuCysAlaAspPr 828
       :::    ::::::|||:::|||    |||    |||
2505 AACCATTATAACCAAACCTCAGGGATGTTCTTGTGCCAGATACCTGGTGT 2554

 828 oTyrGluLeuGlyAspAspGlyArgThrCysValAspSerGlyArgLeuP 845
       ||||||:::          :::|||:::    :::       |||
2555 ATATGAATTC.........GAGTTTTCATGCATAGGGACTCGTAGTTTGG 2595

 845 roArgValThr 848
       ||||||
2596 GCTTTGTGACC 2606
```

seq_name: gb_ov:AF378824

```
seq_documentation_block:
LOCUS       AF378824      2814 bp    mRNA            VRT       16-AUG-2001
DEFINITION  Danio rerio myeloid-specific peroxidase (mpx) mRNA, partial cds;
            alternatively spliced.
ACCESSION   AF378824
VERSION     AF378824.1  GI:15193020
KEYWORDS    .
SOURCE      zebrafish.
  ORGANISM  Danio rerio
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Actinopterygii; Neopterygii; Teleostei; Euteleostei; Ostariophysi;
            Cypriniformes; Cyprinidae; Rasborinae; Danio.
REFERENCE   1  (bases 1 to 2814)
  AUTHORS   Lieschke,G.J., Oates,A.C., Crowhurst,M.O., Ward,A.C. and
            Layton,J.E.
  TITLE     Morphological and functional characterization of granulocytes and
            macrophages in embryonic and adult zebrafish
```

```
  JOURNAL   Blood (2001) In press
REFERENCE   2  (bases 1 to 2814)
  AUTHORS   Crowhurst,M.O., Lieschke,G.J., Layton,J.E. and Ward,A.C.
  TITLE     Direct Submission
  JOURNAL   Submitted (10-MAY-2001) Cytokine Biology Laboratory, Ludwig
            Institute for Cancer Research, The Royal Melbourne Hospital,
            Parkville, Vic 3050, Australia
FEATURES             Location/Qualifiers
     source          1. .2814
                     /organism="Danio rerio"
                     /db_xref="taxon:7955"
                     /chromosome="LGX"
                     /map="between z8146 and z9473"
                     /tissue_type="kidney"
                     /dev_stage="adult"
                     /note="sequence AF326958 deposited in GenBank was used as
                     a probe"
     gene            <1. .2814
                     /gene="mpx"
     CDS             <1. .2037
                     /gene="mpx"
                     /note="myeloperoxidase; similar to mammalian eosinophil
                     peroxidase; alternatively spliced"
                     /codon_start=1
                     /product="myeloid-specific peroxidase"
                     /protein_id="AAK91661.1"
                     /db_xref="GI:15193021"
                     /translation="DYMAQTLRLISEKAHHVHKRSINATNLLTQDELQTITRMTGCEA
                     QTRPPSCRTRPLINKYRTASGVCNNRKNPLLGASNTPFARWLPAQYEDAVSQPKGWDP
                     NKLYNGAALPMVRLVSNRILATADADIESDHDFTFMLTIFGQWVDHDLTFTPFSPSIR
                     SFSNGIDCENSCERSEPCFPISAPPGDQRLRPNTCLPVFRSAPTCGSGHTAYMFGEVP
                     NVREQINTLTAYLDAGQVYGSEDGLAKELRDLTNDGGLLRVNNRFKDNGRELLPFTSV
                     NTNLCATRQKILNDSTLTEVPCFIAGDARVDENPALNSLHTLFVREHNRLARALHVLN
                     PTWSSETLYQEARKIVGAFNQILVIKEYLPLIVGPDAYNRHLGPYPGYNENVDPTIAN
                     VFATAAFRFAHLTIQPFIFRLDENYKNHPQFPSVPLYEAFFSPWRVIFEGGIDPVLRG
                     LIGRPAKLNTQDHMLVNALRERLFAFTSHIALDLASLNMQRGRDHAIPGYNAWRRFCG
                     LSAPQNEQELAVVMNNTELARKLIELYGTPENIDVWLGGVAEPFAPGGRVGSLFACLI
                     SRQFQKIRDGDRLWFESNGVFTTKQKTALASVSMARIICDNTGILKVPRDPFRFRSPA
                     SFVNCGNIPAFDLEPWKETGDGGINISDIGNEIEIPSQDLPDPGDLQENVVHRV"
BASE COUNT      701 a    658 c    618 g    837 t
ORIGIN


alignment_scores:
           Quality: 1619.00              Length:     742
             Ratio:  3.009                 Gaps:      10
 Percent Similarity:  72.507   Percent Identity: 42.588

alignment_block:
 US-08-482-402A-3_COPY_1_848 x AF378824    ..

 Align seg 1/1  to: AF378824  from: 1  to: 2814

     92 GluIleMetGluThrSerIleGlnAlaMetLysArgLysValAsnLeuLy 108
         :::   |||      ::::::::   :::::::::|||::::::
      1 GACTACATGGCACAAACGCTGAGACTGATCAGTGAGAAAGCCCATCATGT 50

    108 sThrGlnGlnSerGlnHisProThrAspAlaLeuSerGluAspLeuLeuS 125
            ::::::|||   :::    |||:::   |||::::::|||   |||:
     51 ACACAAGAGGTCCATCAATGCCACAAACCTGCTCACTCAAGACGAGCTTC 100

    125 erIleIleAlaAsnMetSerGlyCysLeuProTyrMetLeuProProLys 141
         ::   |||::::::|||:::||||||                |||||:::
    101 AAACCATTACACGTATGACAGGCTGTGAAGCTCAAACCCGGCCACCATCT 150

    142 CysProAsnThrCysLeuAlaAsnLysTyrArgProIleThrGlyAlaCy 158
         |||    :::        |||   ||||||||||||         :::|||:::||
    151 TGTAGGACCAGACCCCTCATAAACAAGTATCGAACTGCCAGCGGTGTCTG 200

    158 sAsnAsnArgAspHisProArgTrpGlyAlaSerAsnThrAlaLeuAlaA 175
         |||||||||     :::|||        |||||||||||||||    :::||||
    201 CAACAATCGGAAGAACCCACTTCTTGGTGCATCCAACACTCCATTTGCCC 250

    175 rgTrpLeuProProValTyrGluAspGlyPheSerGlnProArgGlyTrp 191
         |||||||||||      |||||||||:::    |||||||||:::||||||
    251 GCTGGTTGCCTGCTCAATATGAGGACGCCGTTTCTCAGCCAAAAGGATGG 300

    192 AsnProGlyPheLeuTyrAsnGlyPheProLeuProProValArgGluVa 208
         :::|||:::   ||||||||||||      ||||||    ||||||    ||
    301 GATCCCAACAAGTTGTACAATGGTGCAGCGCTGCCCATGGTTAGACTGGT 350

    208 lThrArgHisValIleGlnValSerAsnGluValValThrAspAspAspA 225
         |:::::::::::::::   :::::::::       :::    :::|||
    351 TTCCAATCGCATTCTAGCCACTGCAGATGCTGACATAGAGAGTGACCATG 400

    225 rgTyrSerAspLeuLeuMetAlaTrpGlyGlnTyrIleAspHisAspIle 241
           ::::::   :::|||      :::||||||::::::|||||||||:::
    401 ACTTTACCTTCATGCTCACCATCTTTGGGCAATGGGTTGACCACGATCTG 450

    242 AlaPheThrProGlnSerThrSerLysAlaAlaPheGlyGlyGlySerAs 258
         :::|||||||||   |||   |||      :::|||:::::|||    ||
    451 ACTTTTACACCTTTCTCACCAAGTATTAGGTCCTTTAGTAATGGCATTGA 500

    258 pCysGlnMetThrCysGluAsnGlnAsnProCysPheProIleGlnLeuP 275
         ||||:::    :::||||||:::::::::||||||||||||||||:::    |
    501 CTGTGAAAACAGTTGTGAGCGCTCTGAACCCTGCTTCCCAATTTCGGCCC 550

    275 roGluGluAlaArgProAlaAlaGlyThrAlaCysLeuProPheTyrArg 291
         ||           :::          ::::::|||||||||    :::|||
    551 CTCCAGGAGATCAGCGTCTGCGCCCTAACACCTGTCTCCCTGTCTTCCGG 600

    292 SerSerAlaAlaCysGlyThrGlyAspGlnGlyAlaLeuPheGlyAsnLe 308
         |||:::    :::||||||:::|||        :::    :::||||||:::::
    601 TCAGCACCGACTTGTGGTTCTGGGCACACTGCCTACATGTTTGGTGAAGT 650

    308 uSerThrAlaAsnProArgGlnGlnMetAsnGlyLeuThrSerPheLeuA 325
         :         |||   |||:::|||:::|||   ||||||::::::||||
    651 C......CCAAATGTTCGGGAACAGATCAATACTCTAACAGCTTACCTAG 694

    325 spAlaSerThrValTyrGlySerSerProAlaLeuGluArgGlnLeuArg 341
         ||||||:::     ||||||||||||:::       :::|||      ::::::||||||
    695 ATGCTGGACAAGTTTACGGTTCAGAGGATGGACTAGCAAAGGAACTGCGG 744

    342 AsnTrpThrSerAlaGluGlyLeuLeuArgValHisGlyArgLeuArgAs 358
         :::    |||:::        ||||||||||||||||:::::|||::::::||
    745 GACCTTACTAATGATGGTGGCCTTTTACGTGTTAATAATCGTTTTAAAGA 794

    358 pSerGlyArgAlaTyrLeuProPheValProProArgAlaProAlaAlaC 375
         |:::||||||       |||||||||:::         ::::::          |
    795 CAATGGACGAGAGCTCCTGCCCTTTACTAGTGTTAACACC...AACTTGT 841

    375 ysAlaProGluPro.........GlyAsnProGlyGluThrArgGlyPro 388
         |||||    :::               ::::::          |||:::    |||
    842 GCGCCACTCGTCAGAAAATCCTCAACGACAGCACTCTCACAGAGGTTCCT 891

    389 CysPheLeuAlaGlyAspGlyArgAlaSerGluValProSerLeuThrAl 405
         ||||||:::|||||||||:::|||::::::|||      |||:::|||:::::
    892 TGCTTTATTGCAGGTGATGCTCGTGTCGATGAGAATCCTGCTCTGAACTC 941

    405 aLeuHisThrLeuTrpLeuArgGluHisAsnArgLeuAlaAlaAlaLeuL 422
         :|||||||||||::::::||||||||||||||||||||      ||||||
    942 CTTACACACACTTTTTGTAAGGGAGCACAATAGATTGGCTCGCGCCCTTC 991

    422 ysAlaLeuAsnAlaHisTrpSerAlaAspAlaValTyrGlnGluAlaArg 438
           :::||||||       ||||||:::::::::::||||||||||||||||
    992 ATGTGCTCAACCCTACCTGGAGTAGTGAAACTCTTTATCAAGAAGCAAGA 1041

    439 LysValValGlyAlaLeuHisGlnIleIleThrLeuArgAspTyrIlePr 455
         |||:::|||||||||::::::||||||:::::::::::::::|||:::||
   1042 AAAATTGTCGGTGCATTCAACCAGATCTTGGTGATCAAGGAATACTTGCC 1091

    455 oArgIleLeuGlyProGluAlaPheGlnGlnTyrValGlyProTyrGluG 472
         |    |||:::||||||:::|||::::::::::::::|||||||||    |
   1092 ACTAATTGTAGGCCCTGATGCTTACAATAGACACCTTGGACCATATCCAG 1141

    472 lyTyrAspSerThrAlaAsnProThrValSerAsnValPheSerThrAla 488
```

```
          |||||::::::::::::::||||||:::::||||||||||:::||||||
1142 GTTACAATGAGAATGTAGACCCTACCATTGCGAACGTCTTTGCCACTGCT 1191

 489 AlaPheArgPheGlyHisAlaThrIleHisProLeuValArgArgLeuAs 505
     ||||||||||||:::|||    |||||:::|||::::::   ||||||||
1192 GCTTTCCGTTTGCCCACTTGACTATCCAGCCCTTCATTTTTCGTCTTGA 1241

 505 pAlaSerPheGlnGluHisProAspLeuProGlyLeuTrpLeuHisGlnA 522
     |    ::::::::::::||||||::::::|||::::::   |||::::::|
1242 TGAAAACTACAAGAACCATCCTCAGTTTCCCAGTGTCCCCCTCTATGAGG 1291

 522 laPhePheSerProTrpThrLeuLeuArgGlyGlyGlyLeuAspProLeu 538
     ||||||||||||||||     ::::::         ||||||:::||||||:::
1292 CTTTCTTCTCTCCATGGAGGGTGATCTTTGAAGGGGGAATTGATCCTGTG 1341

 539 IleArgGlyLeuLeuAlaArgProAlaLysLeuGlnValGlnAspGlnLe 555
     :::||||||||||::::::||||||||||||||||||::::::||||||:::::
1342 CTCCGTGGATTGATTGGTCGTCCGGCAAAACTGAACACCCAGGACCACAT 1391

 555 uMetAsnGluGluLeuThrGluArgLeuPheValLeuSerAsnSerSerT 572
     ::::   :::    |||     |||||||||||||::::::::::::         :
1392 GTTGGTGAATGCTCTCAGAGAGAGGCTGTTTGCCTTCACATCCCACATAG 1441

 572 hrLeuAspLeuAlaSerIleAsnLeuGlnArgGlyArgAspHisGlyLeu 588
     ::|||||||||||||||:::|||:::||||||||||||||||||||:::::
1442 CTTTGGATCTGGCTTCCCTCAACATGCAAAGAGGGCGTGACCATGCTATA 1491

 589 ProGlyTyrAsnGluTrpArgGluPheCysGlyLeuProArgLeuGluTh 605
     ||||||||||||    ||||||:::||||||||||||          :::::
1492 CCAGGTTATAATGCATGGCGTCGGTTCTGTGGATTGTCTGCTCCTCAGAA 1541

 605 rProAlaAspLeuSerThrAlaIleAlaSerArgSerValAlaAspLysI 622
     :        :::|||::::::::::::    :::     ::::::|||    |||:
1542 TGAGCAAGAATTGGCTGTGGTGATGAACAACACTGAACTAGCCCGCAAGT 1591

 622 leLeuAspLeuTyrLysHisProAspAsnIleAspValTrpLeuGlyGly 638
     :::::::||||||         |||:::||||||||||||||||||||||||
1592 TAATTGAGCTTTATGGCACCCCTGAGAACATTGATGTTTGGTTAGGAGGT 1641

 639 LeuAlaGluAsnPheLeuProArgAlaArgThrGlyProLeuPheAlaCy 655
     :::|||||||    |||    |||    :::||||:::|||    |||||||||||
1642 GTTGCTGAGCCTTTTGCTCCTGGTGGTCGTGTCGGTTCTCTTTTTGCCTG 1691

 655 sLeuIleGlyLysGlnMetLysAlaLeuArgAspGlyAspTrpPheTrpT 672
     |||||||::::::|||::::::    :::||||||||||||    :::|||:
1692 CCTGATTTCAAGACAGTTCCAGAAAATCCGAGATGGCGATAGGTTGTGGT 1741

 672 rpGluAsnSerHisValPheThrAspAlaGlnArgArgGluLeuGluLys 688
     ::|||::::::    |||||||||       |||:::       |||     :::
1742 TTGAGAGCAATGGAGTGTTCACCACCAAACAGAAAACAGCTCTGGCTTCT 1791

 689 HisSerLeuSerArgValIleCysAspAsnThrGlyLeuThrArgValPr 705
        |||::::::|||:::||||||||||||||||||:::     :::||||||
1792 GTGTCAATGGCCCGCATAATCTGCGACAACACCGGAATCCTTAAAGTTCC 1841

 705 oMetAspAlaPheGlnValGlyLysPheProGluAspPheGluSerCysA 722
     |    |||   |||:::        :::    |||     :::|||     :::|||
1842 CAGAGATCCTTTTCGCTTC...CGGAGTCCTGCCAGTTTTGTTAACTGTG 1888

 722 spSerIleThrGlyMetAsnLeuGluAlaTrpArgGluThr......... 735
       :::|||    :::::::::||||||    |||:::||||||
1889 GAAATATCCCAGCTTTTGATCTCGAGCCTTGGAAAGAGACTGGTGACGGT 1938

 736 ..................................PheProGlnAspAs 740
                                       :::|||::::::||
1939 GGCATCAACATTTCAGACATTGGAAATGAGATAGAGATTCCATCCCAGGA 1988

 740 pLysCysGlyPheProGluSerValGluAsnGlyAspPheValHisCysG 757
     |              :::         ||||||         ||||||
1989 CCTCCCGGACCCAGGGGACCTCCAGGAGAAC......GTGGTCCACAGGG 2032

 757 luGluSerGlyArgArgValLeuValTyrSerCysArgHisGlyTyrGlu 773
          :::|||::::::                                   :::

2033 TGTAGCAGGGCCACCAG............................GAC 2052

 774 LeuGlnGlyArgGluGlnLeuThrCysThrGlnGluGlyTrpAspPheGl 790
     ||||||         ::::::|||           |||:::        |||     ::
2053 CTCCAGGAATTCCAGGACCTCCCATTAATACAACAG......GACAACAG 2096

 790 nProPro.LeuCysLysAspValAsnGluCysAlaAspGlyAlaHisPro 806
     :    |||  |||                        |||     :::       ||||||
2097 TCTGCCTTCTT....................TGCCTCAGTCAACTCCATCCT 2128

 807 ProCysHisAlaSerAlaArgCys 814
        ||||||:::     :::    |||
2129 TCCTGCCACTGCTAAAGTTGTTGT 2152
```